AUTOMATED EXAMINATION OF RELATIVE AFFERENT PUPILLARY DEFECT USING PUPIL DETECTION AND TRACKING

A Dissertation
Presented to
The Academic Faculty

By

Melvin J. Mathew

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in the
School of Electrical and Computer Engineering

Georgia Institute of Technology

August 2018

AUTOMATED EXAMINATION OF RELATIVE AFFERENT PUPILLARY DEFECT USING PUPIL DETECTION AND TRACKING

Approved by:

Dr. G. AlRegib, Advisor
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. E. Moore
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Y. M. Khalifa
School of Medicine
Emory University

Dr. D. Temel
School of Electrical and Computer Engineering
Georgia Institute of Technology

Date Approved: May TBD, 2018

ACKNOWLEDGEMENTS

First and foremost, I would like to thank Prof. Ghassan AlRegib for giving me the opportunity to work on this project under his advisement. This experience has given me exposure to interesting and innovative projects being conducted in the OLIVES lab by some of the brightest and hard-working individuals I have had the honor to learn from. I would also like to thank Dr. Yousuf Khalifa for not only serving on my thesis reading committee, but also for aiding in the collection of data pertinent to this thesis. I also thank Prof. Elliot Moore for serving on my thesis reading committee. I thank Dogancan "Can" Temel for his close support and immeasurable patience throughout the duration of this thesis. Can has not only become a great mentor for me, but he has also become a great friend, always eager to share his tasty Turkish treats!

I would like to thank my parents, family and relatives for all their love and support throughout this journey, allowing me to pursue my ambitions while traveling great distances to get my education, without needing to worry about leaving home. Last and certainly not least, I want to thank all of those countless people who have provided support in a multitude of ways during my studies, all the way from my hometown Gaborone, Botswana, to those from my time spent at the University of Cape Town, South Africa, to the amazing staff and classmates from Georgia Tech Shenzhen, China, and finally to the many I have met at Georgia Tech Atlanta, USA. The journey thus far has been very exciting, and it is surely only just beginning!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

In traditional clinical practice, medical professionals typically examine ocular conditions by conducting a series of procedures using various medical equipment, and these examinations are subject to the environment in which the medical examination takes place. The variability in testing setup can potentially lead to inconsistency in medical assessments [1], however, supposing that these factors can be perfectly controlled, there is still a factor of inconsistency in the ocular condition assessment. Assessments are typically performed visually, which depends on the professional's medical experience, knowledge, and personal health. More so, there is an increasing presence of ocular conditions globally [2], with a shortage and even stagnant presence of eye care professionals [3, 4]. Therefore, there is an increasing need for more consistent, accurate and automated approach to medical examination of ocular conditions.

Swinging flashlight test is a medical procedure used to determine Relative Afferent Pupillary Defect (RAPD) [5], a common indicator of ocular disease. RAPD is the condition in which the pupillary light reflex response of both pupils differ to one another. The severity of RAPD is correlated to the similarity or dissimilarity of the pupillary light reflex responses [6]. In this thesis, we propose automating RAPD examination using pupil detection and tracking algorithms to measure the pupil response during the swinging flashlight test. Specifically, we utilize image pre-processing techniques including temporal cropping, downsampling, and spatial cropping, to improve pupil detection and pupil size measurement by the Circular Hough Transform. RAPD severity score is obtained by performing Pearson Correlation Coefficient on the measured pupil responses, which quantifies the similarity or dissimilarity. We also perform post-processing techniques on the obtained pupil size measurements, including median filtering, thresholding, and moving average, to minimize noise and incorrect pupil detection measurements. The proposed algorithmic pipeline achieves 84% accuracy on the available testing set.

CHAPTER 1

INTRODUCTION

The American Academy of Ophthalmology (AAO) report statistics from a survey conducted in 2010, which showed that among Americans, more than 35 million instances of eye diseases were recorded [2]. We summarize these statistics in Fig. 1, indicating the significant presence of serious ocular conditions such as cataract, age-related macular degeneration (AMD), and diabetic retinopathy. As the numbers continue to grow, the statistics from 2010 are estimated to double by 2050. These eye conditions have to be dealt with by eye care specialists known as ophthalmologists, however, the AAO reported less than 20 thousand active ophthalmologists in the United States [2]. A more recent article published by Ophthalmology Management in 2015 [3] reports an increasing demand for eye care specialists, however, the article also reports a lack of federal funding for Ophthalmology. Therefore, with an increasing presence of ocular conditions and a decreasing influx of eye care specialists, there is a need to automate medical examination of ocular diseases.

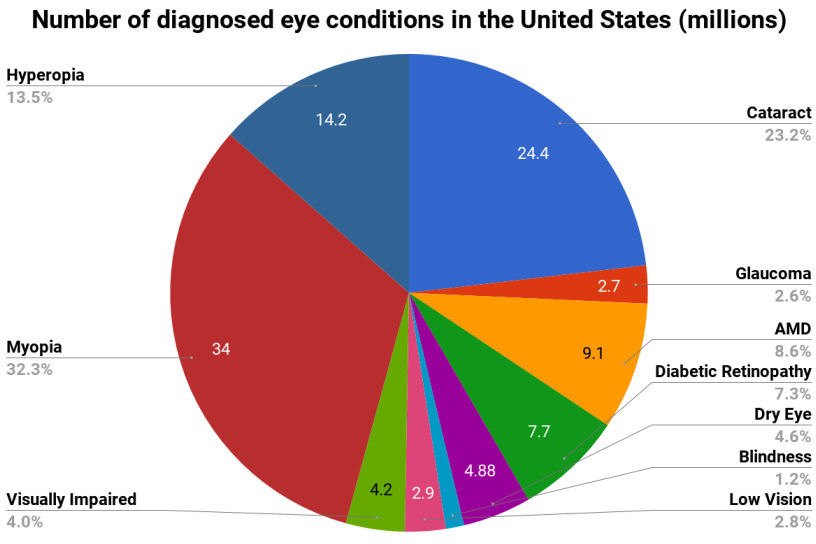


Figure 1: Number of diagnosed eye conditions in the United States (millions).

In addition to the shortage of specialists, there is also inconsistency in traditional ocular examination, as reported in a study by Omburo et al. [1]. The study showed that when using a standardized light source for pupillary assessment, post-anaesthesia care unit nurses were able to make more consistent assessments, when compared to using various light sources. The environment in which the examination take place also affects the medical assessment. For example, the measurement of pupil dilation by light stimulation requires a "dark" room for which to carry out the test. However, "dark" here can be subjective, and often cannot be guaranteed consistent among different medical examination rooms, even within the same hospital or clinic. Therefore, there is a need for an objective, repeatable and accurate approach to medical examination of ocular disease.

A common clinical procedure for examining oculuar disease is known as the swinging flashlight test. This is used to detect a condition known as Relative Afferent Pupillary Defect (RAPD), which is an indicator of ocular disease capable of damaging the optic nerve. The swinging flashlight test is performed by shining a light repeatedly back and forth from one eye to the other, and observing the pupillary light reflex response upon light stimulation. This test is a visual assessment of the pupillary light reflex, in which the dissimilarity between both pupils is estimated by the eye care professional. The greater the dissimilar-

Figure 2: Patient to RAPD score pipeline in practice.

ity (or the lesser the similarity) between both pupils' pupillary light reflex responses, the greater is the severity of RAPD. This process is time consuming and relies on the observer's ability to adequately assess the extent of RAPD present in the patient. Therefore, we automate this process by understanding and interpreting the different stages of this clinical procedure, from an algorithmic point of view. In Fig. 2 we illustrate the clinical pipeline in which patients are mapped to assessment scores by the process of physical examination by a specialist.

The physical process of the swinging flashlight test can be automated using hardware, utilizing light-emitting diodes (LEDs) to serve as light stimuli, and cameras to record videos of the pupillary light reflex responses. We can automate the visual assessment conducted by the eye care professionals using pupil detection and tracking algorithms, for pupil size estimation. By utilizing a dissimilarity or similarity measure on the pupil size measurements, and by determining the reactivity of the pupil upon light stimulation, we can obtain the corresponding RAPD severity score. Reactivity upon light stimulation refers to the amount of pupil size dilation when a light has been shone onto the pupil. If a pupil has no reactivity or minimal dilation relative to the adjacent pupil, and if the pupil responses have a high correlation, there is an indication of severe RAPD.

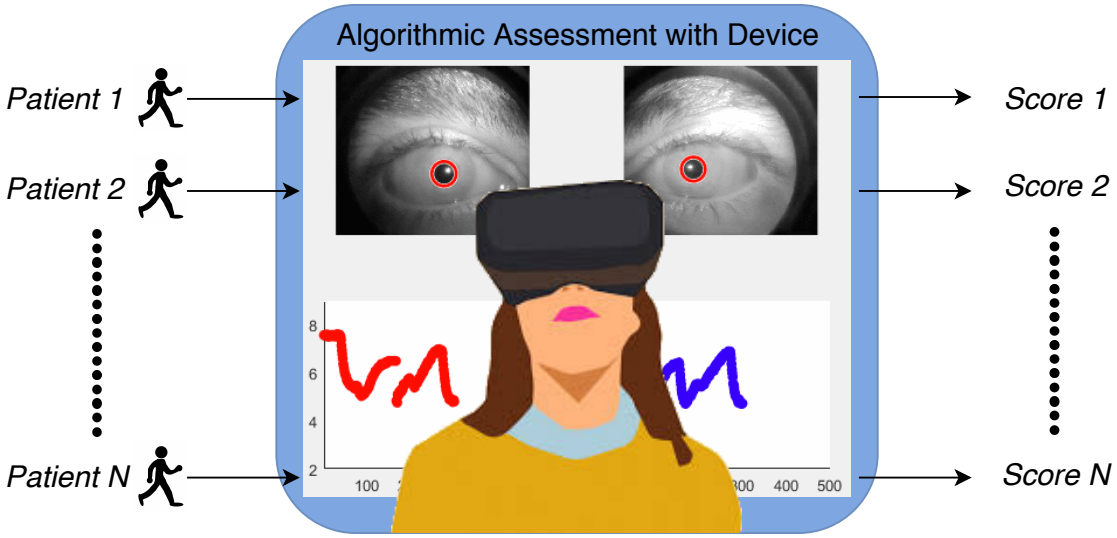


Figure 3: Algorithmic pipeline for patient to RAPD score.

Therefore, in this thesis, the proposed pipeline for automated examination of RAPD from an algorithmic point of view can be interpreted as mapping patient video recordings of pupillary light reflex responses to similarity scores, which are the algorithmic RAPD scores. In Fig. 3, we illustrate the algorithmic equivalent of the same clinical pipeline described previously. The inputs are patients and the outputs are RAPD severity scores, however, the process here includes algorithmic components for automated assessment, rather than visual examination by a specialist.

The algorithmic components presented in this thesis include pre-processing for pupil image enhancement, Circular Hough Transform for pupil detection and pupil size measurement, and post-processing for pupil size measurement noise reduction. Pre-processing stages include temporal cropping, spatial downsampling, and spatial cropping. Circular Hough Transform includes Canny edge detection, Sobel operation, Hough transformation, and voting for pupil contour circle fitting. Post-processing techniques include median filtering, thresholding, and moving average.

CHAPTER 2

LITERATURE REVIEW

In this chapter, we briefly describe Relative Afferent Pupillary Defect (RAPD) and discuss the examination of this condition in clinical practice. We then present the literature of pupil detection and tracking algorithms for pupil size measurement, and describe the relation of this thesis to the literature. We divide the literature into three main sections, namely, examination of relative afferent pupillary defect in clinical practice, image pre-processing for pupil image enhancement, and pupil contour detection for pupil size measurement.

2.1 Examination of Relative Afferent Pupillary Defect in Clinical Practice

Relative Afferent Pupillary Defect (RAPD), otherwise known as Marcus Gunn Pupil [7], is the condition in which the reaction of both pupils to a light stimulus differs from one

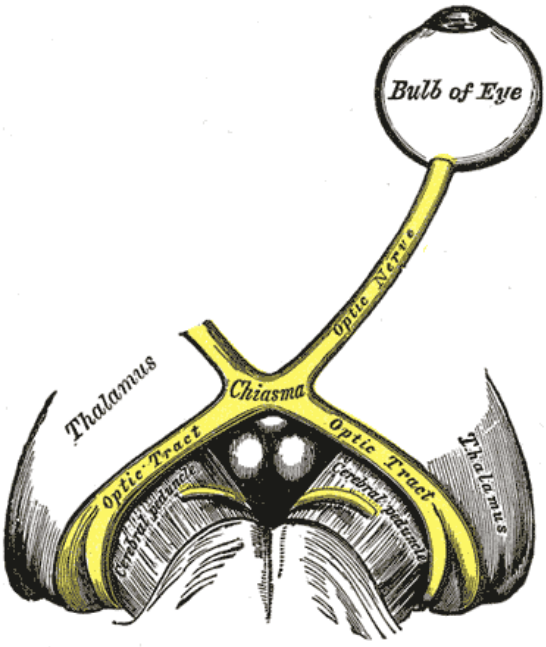


Figure 4: The left optic nerve and the optic tracts [7].

another. This condition can occur when the optic nerve is damaged, indicating the presence of ocular disease [5] such as optic neuritis, ischemic retinal disease, severe glaucoma causing trauma to optic nerve, direct optic nerve damage (trauma, radiation, tumor), retinal detachment, very severe macular degeneration, and retinal infection (CMV, herpes). The optic nerve links the eye to the optic tracts as illustrated in Fig. 4.

Relative Afferent Pupillary Defect (RAPD) is visually assessed, and is seen in the direct response. The direct response refers to the constriction of the pupil which is exposed to a light stimulus. A normal pupillary light reflex to a light stimulus is one in which the pupil constricts upon stimulation, and the adjacent pupil mirrors the effect of being stimulated. However, if RAPD is present in a pupil, when the light is moved from the unaffected eye to the affected eye, both pupils will appear to dilate, as if the affected eye does not respond to the light stimulus. Therefore, there is an indication of RAPD when both pupils do not have similar direct responses.

There are different levels of RAPD severity, typically scored in integers between zero and four. Manual RAPD examination is dependent on the observer's ability to adequately assess the severity of RAPD, which suggests the clinical practice may be a subjective process. A common test used to assess RAPD and its severity is known as the swinging flashlight test.

2.1.1 Swinging Flashlight Test

Swinging flashlight test is a common clinical procedure for diagnosing Relative Afferent Pupillary Defect (RAPD) [5, 6]. An observer swings a light between both pupils and visually compares any constriction in both pupils. Fig. 5 illustrates this procedure, in

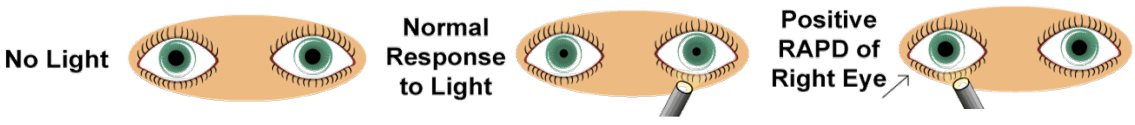


Figure 5: Swinging flashlight test indicating positive test for RAPD [5].

which the pupils are shown before and after testing. When no light is shone onto the subject's eyes, the pupils are at maximum dilation size. When a healthy eye is stimulated by a light source, the pupillary light reflex response is triggered in the adjacent eye, and both eyes will constrict equally.

However, for an eye containing Relative Afferent Pupillary Defect (RAPD), the pupil being stimulated may not constrict fully, and for severe cases, the pupil may not constrict at all. If the adjacent eye does not have this defect, the adjacent pupil's pupillary light reflex response will reflect the same level of constriction as the affected eye. However, if the adjacent eye constricts fully, then the stimulated eye may have some other defect other than RAPD, such as a muscle condition that prevents pupil constriction or dilation. Therefore, adequate assessment of RAPD requires examination of pupil reactivity to light stimulation, as well as examination of pupillary light reflex response in the non-stimulated pupil.

2.2 Image Pre-processing for Pupil Image Enhancement

Pupil can be algorithmically detected based on the physical characteristics of the pupil, such as its circular shape and black color. However, other regions within the eye may share similar properties to the pupil, such as circular iris, or black colored eyebrows. Therefore, this section discusses literature for image processing techniques used in enhancement of the pupil within the image, with respect to the other regions of the eye. The purpose of pupil image enhancement is to improve algorithmic pupil detection, by making the pupil algorithmically and visually distinct from the other regions of the eye.

Pre-processing techniques vary depending on the input image conditions. This thesis handles data obtained from a device that captures pupil data under specified, controlled conditions, in which only the eye regions of the subjects are captured. Therefore, the literature discussed in this thesis will not involve region of interest extraction techniques for eye localization, commonly seen in pupil detection literature.

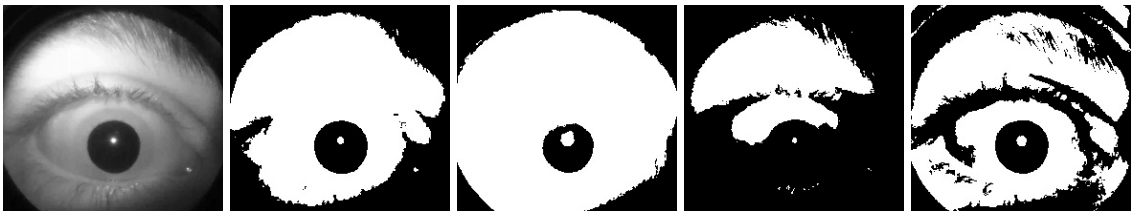
State-of-the-art pupil detection and tracking algorithms typically utilize image pre-processing techniques such as Binary Threshold [8, 9], Canny [9–12], Color Space [9], and Sobel [8, 9, 12, 13]. These methods primarily focus on the edge characteristics of the pupil to enhance the edge distinction between the pupil and the other regions of the eye. We briefly discuss these techniques in the following subsections.

2.2.1 Binary Threshold

Binary threshold is used to binarize the image by comparing the image pixel values to a set threshold value [14]. Pixels above or below this threshold are set to black and white accordingly, in which darker regions become black (minimum pixel value) and brighter regions become white (maximum pixel value). In the context of pupil image enhancement, binary threshold is used to create a clearer distinction between the pupil (darker regions) and the other regions of the eye such as the iris and surrounding skin (brighter regions).

$$I(x,y) = \begin{cases} 0 & if I(x,y) > T(x,y) \\ 255 & otherwise \end{cases} \tag{1}$$

Threshold value is image dependent, however, it can be determined adaptively. For example, a pixel located at (x, y) in an 8-bit image I, can be binarized with a threshold value that is dependent on some function of the surrounding neighborhood pixels at that



(a) **(b)** **(c)** **(d)** **(e)**

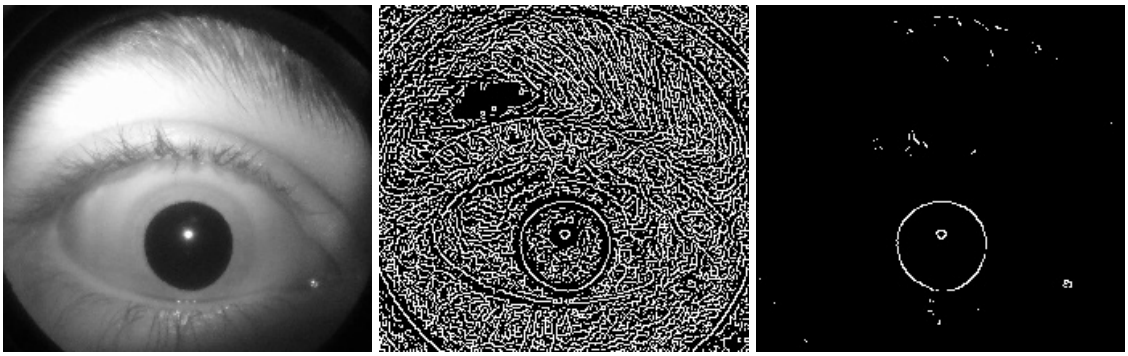
Figure 6: (a) Grayscale image (230×216 pixels). (b) Binary image after thresholding, $T(x, y) = 100$. (c) Binary image after thresholding, $T(x, y) = 50$. (d) Binary image after thresholding, $T(x, y) = 150$. (e) Binary image after adaptive thresholding, $T(x, y) = 51 \times 51$ neighborhood mean.

image pixel point. Adaptive threshold values can be determined by taking the mean or weighted sum [15] of the neighbordhood pixels to serve as the threshold value for that specific image pixel. The binary operation is represented by Equation 1, where $T(x, y)$ is the threshold value function for the binary operation applied to the image pixel at image coordinate (x, y).

We illustrate case examples of various binary threshold operations in Fig. 6, in which a sample image is binarized using different thresholds. In Fig. 6c we see that a constant threshold value can isolate the pupil from the other regions of the eye. However, this non-adaptive threshold requires manual tuning of the threshold value, which may vary from image to image as illustrated by choosing a different threshold value in Fig. 6b and Fig. 6d. This is unlike the adaptive threshold seen in Fig.6e, which selects a threshold based on the neighborhood mean at every pixel location.

2.2.2 Canny

Canny edge detection is a five-step algorithm used to find edges within an image. These steps include gaussian smoothing, image gradient calculation, non-maximum suppression, double threshold, and edge tracking using hysteresis, in which weak disconnected edges are suppressed. We illustrate sample Canny edge detection results in Fig. 7, using a sample



(a) **(b)** **(c)**

Figure 7: (a) Grayscale input image (230×216 pixel). (b) Canny edges, threshold 0 (a). (c) Canny edges, threshold 255.

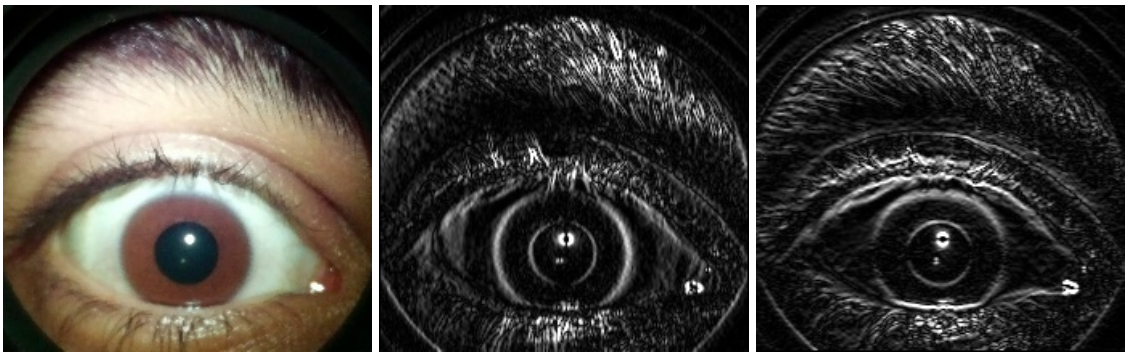
grayscale image. In Fig. 7b, we observe the result of determining all edges in the image by applying no thresholds or constraints, while in Fig. 7c we observe the result of constraining detected edges to only include those with the steepest gradients within local regions of the image.

It is evident from the sample images shown in Fig. 7 that the Canny edge detector is sensitive to the thresholds and constraints provided. The result seen in Fig. 7b will not be useful for pupil detection, however, a result such as that shown in Fig. 7c is able to effectively isolate the pupil. Therefore, the threshold values selected require manual tuning based on the image properties of the data being handled.

2.2.3 Sobel

Sobel operator (or Sobel filter) is an edge detection algorithm in which the Sobel operator is convolved with the image such that the edges of the image are emphasized. The Sobel operator determines edges by determining the points in an image which have a high gradient in a particular direction, while having a low gradient in the orthogonal direction. We illustrate the effect of performing the Sobel operation on a sample eye image in Fig. 8.

In Fig. 8b, we use a Sobel operator that emphasizes the edges of the color image shown in Fig. 8a, in which the operator calculates gradients along the x-axis direction of the



(a) **(b)** **(c)**

Figure 8: (a) Original color input image (230×216 pixels). (b) 3×3 Sobel applied to (a) in x-axis direction. (c) 3×3 Sobel applied to (a) in y-axis direction.

image. Similarly, we do the same for the y-axis direction of the image in Fig. 8c. In both these examples, we notice the operator is capable of highlighting the edges around the iris as well as the pupil. This is possible due to the contrast in pixel intensities between the white regions of the eye and the iris, and between the iris and the pupil.

2.2.4 Color Space

Color images are commonly represented using the Red-Blue-Green (RGB) color space. RGB image contains three channels, one for each color dimension. Each channel of the three-channel RGB image contains the intensity values of the respective color channel. However, we can also represent color images using the Lab color space, in which L refers

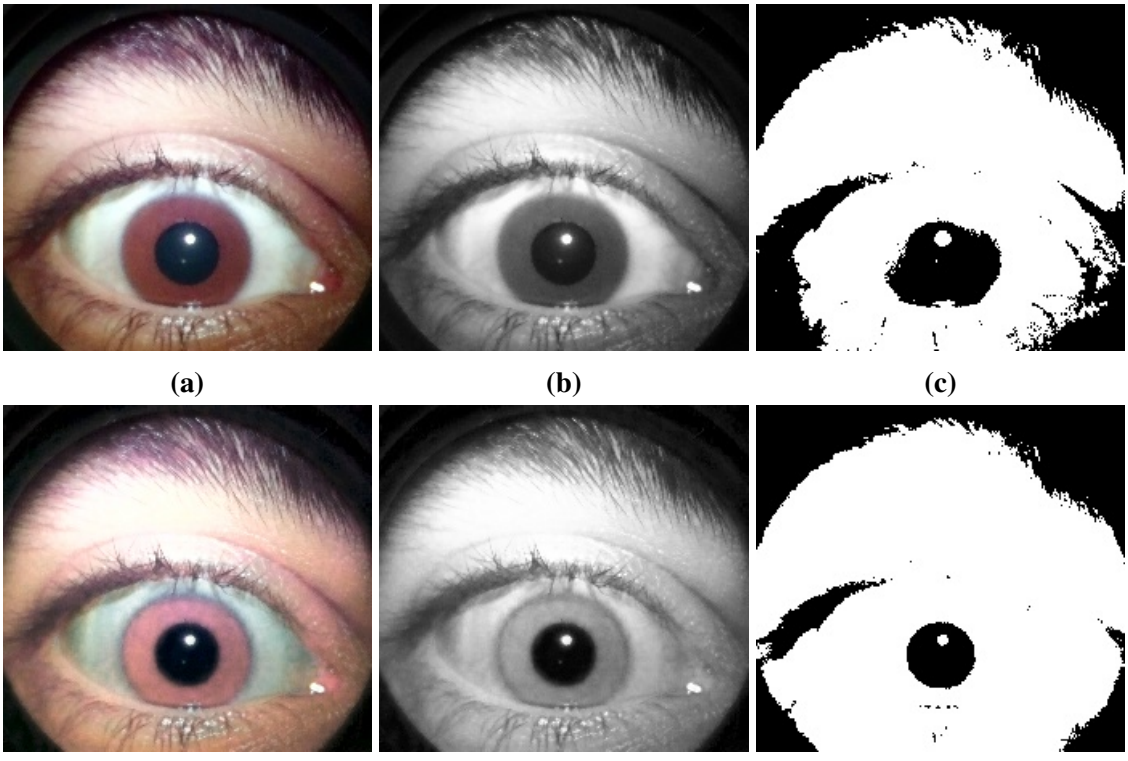


Figure 9: (a) Original RGB color image (3-channel, 230×216 pixel per channel input RGB image). (b) Grayscale of original image (a). (c) Adaptive binary operation using mean of grayscale image (b) as threshold. (d) Enhanced pupil color image of (a) after color space manipulation, using RGB and Lab color space representations of (a). (e) Grayscale of color manipulated image (d). (f) Adaptive binary operation using mean of grayscale image (e) as threshold.

to lightness, and a and b refer to the color-opponent dimensions. Lab color space covers all possible perceivable colors and is independent of the device in which the image is displayed. Therefore, we can effectively manipulate the image channels in both RGB and Lab space to achieve a greater distinction between the pupil and the other regions of the eye [9].

An example of color space manipulation for pupil image enhancement is illustrated in Fig. 9. We notice that the color space manipulated image in Fig. 9d results in an improved pupil isolation, emphasized by the binarized image in Fig. 9f. This is contrasted with the binarized original image in Fig. 9a through Fig. 9c, in which we observe the pupil is not easily distinguishable from the iris.

2.3 Pupil Contour Detection for Pupil Size Measurement

State-of-the-art pupil detection and tracking algorithms utilize the pre-processing techniques discussed in the previous sections. Adachi et al. [9] present a color space manipulation approach to pupil image enhancement for improved pupil detection using circle fitting, for images under natural lighting conditions. Masi et al. [8] utilize a binary threshold approach and Sobel operation for pupil extraction. The work presented by Swirski et al. [16] proposes a haar-feature based approach to pupil radius estimation, utilizing K-means clustering, thresholding, and ellipse fitting using random sample consensus (RANSAC). Pupil labs [17] proposes a Canny based approach with ellipse fitting. SET [18] incorporates a threshold based segmentation approach to pupil extraction. ExCuSe [19] utilizes morphological operations for pupil image enhancement [16, 18] and relies on edge characteristics to determine the pupil's position. ElSe [20] is an edge based approach for pupil detection. These state-of-the-art algorithms typically utilize some variation of the Circular Hough Transform [21, 22] or the Starburst algorithm [16, 18, 23] for pupil contour detection and size estimation, either by circle or ellipse fitting [20] respectively.

2.3.1 Circular Hough Transform

The Circular Hough Transform (CHT) is a variation of the Hough transformation for line detection, used to find circles in an image. CHT maps a 2-dimensional point in the image space to a right circular cone in the Hough space [24]. Circles are found in an image by mapping all edge points to the Hough space and determining the most significant points of intersection of the right circular cones. These points of intersections correspond to circle parameters in the image space. Voting is used to determine significant points of intersection, and by constraining parameters such as number of circles, minimum radius, and maximum radius, the most relevant circle parameters are returned. The edge points used in the Hough transformation are found using Canny edge detector and Sobel operator. CHT is illustrated in Fig. 10.

(a) **(b)** **(c)** **(d)** **(e)**

Figure 10: (a) The original image. (b) The detected Canny edge point. (c) The gradient image after Sobel operator in x-direction. (d) The gradient image after Sobel operator in y-direction. (e) The detected circle.

The Circular Hough Transform compares the detected edge points shown in Fig. 10b, of the original image in Fig. 10a, with the corresponding gradient magnitudes in both x-axis and y-axis directions, using the gradient images shown in Fig. 10c and Fig. 10d. The gradient magnitudes are utilized to constrain points such that the circle center is normal to the selected Canny edge points. These points are transformed into the Hough space, in which every point corresponds to a right circular cone. The point of intersection of all these cones refers to the circle's parameters, which is illustrated in Fig. 10e.

2.3.2 Starburst Algorithm

Starburst Algorithm is primarily used in gaze-tracking applications [23], however, the techniques used in Starburst can be applied to general pupil detection and measurement [25]. A feature-based pupil contour is obtained using this algorithm, in which the contour points are found by taking derivatives along projected rays from a best-guess pupil center coordinate [23]. The pupil's diameter is estimated from the obtained contour feature points. Starburst algorithm implements ellipse fitting algorithm to best describe the contour [8] by utilizing the ellipse equation to map feature points to a best-fit ellipse [26].

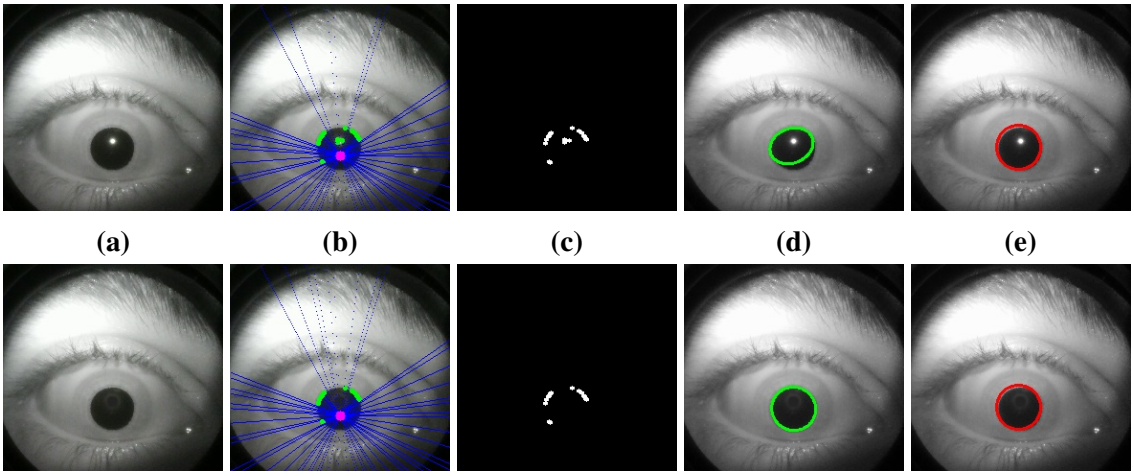


Figure 11: (a) The original image. (b) The Starburst algorithm after a single iteration of finding feature points using projected rays, using a gradient threshold of 75 pixel intensity. (c) The feature points detected. (d) The result of best-fit ellipse on feature points shown in (c). (e) The result of Circular Hough Transform on feature points shown in (c). (f) – (j) The algorithm performed after corneal reflection removal.

In Fig. 11 we illustrate the sensitivity of the ellipse equation to incorrect feature points. We see that the corneal reflection within the pupil is detected as feature points in Fig. 11c, which is due to high gradient magnitudes when moving from the reflection to outside the reflection within the pupil. These feature points do not belong to the pupil's contour, and the corresponding best-fit ellipse using all these feature points is shown in Fig. 11d. We observe that ellipse fitting with unconstrainted feature points leads to an incorrect pupil

measurement when contrasted with the Hough Circular Transform result in Fig. 11e, which is more robust to incorrect feature points.

However, Starburst algorithm addresses this limitation by adding a pre-processing stage to eliminate corneal reflection. This can be done by utilizing the brightness of the corneal reflection. Therefore, the Starburst algorithm is initially used to identify the corneal reflection and remove it using Gaussian smoothing, after which the Starburst algorithm can be applied for pupil detection. The result after corneal reflection removal is illustrated in Fig. 11f through Fig. 11j. Starburst algorithm also applies a random sample consensus (RANSAC) model to determine the feature points that best-fit an ellipse.

2.4 Summary

In this section we summarize the literature of state-of-the-art algorithms for pupil detection and tracking, and describe the relation to this thesis. We summarize the techniques utilized by these state-of-the-art algorithms in Table 7.

Table 1: State-of-the-art pupil detection and tracking algorithms.

Algorithms	Adachi	Masi	Starburst	Swirski	Pupil Labs	SET	ExCuSe	ElSe
Binary Threshold	✓	✓	✓	✓	✗	✓	✗	✗
Canny Edges	✗	✗	✗	✓	✓	✗	✓	✓
Morphological Operators	✓	✗	✓	✓	✗	✗	✓	✓
Sobel Operator	✗	✓	✗	✗	✗	✗	✗	✗
Ellipse Fitting	✗	✗	✓	✓	✓	✓	✓	✓
Circle Fitting	✓	✓	✗	✗	✗	✗	✗	✗
Color Space	✓	✗	✗	✗	✗	✗	✗	✗
Temporal Characteristics	✗	✗	✗	✗	✗	✗	✗	✗

2.4.1 Image Pre-processing for Pupil Image Enhancement

The pre-processing techniques discussed in this chapter utilize the edge characteristics of the pupil, by applying binary thresholding, canny edge detection, and Sobel operation, for pupil isolation or extraction. These techniques were shown to be dependent on parameters that are typically tuned manually, dependent on the application. However, adaptive thresholds utilize image information such as the neighborhood mean to automate the parameter tuning. Color characteristic of the pupil has also been used to create a greater distinction between the pupil and the other regions of the eye. However, none of the described methods utilize temporal characteristics of the pupil data.

2.4.2 Pupil Contour Detection for Pupil Size Measurement

State-of-the-art pupil detection algorithms utilize the Circular Hough Transform or the Starburst Algorithm for pupil contour fitting, either by circle or ellipse fitting respectively. These algorithms utilize edge based detectors, gradient information, and morphological operations to identify the pupil within the image. However, color and temporal characteristics are typically not used. The Starburst algorithm utilizes post-processing stages to filter incorrect feature points for ellipse fitting, using random sample consensus (RANSAC). However, the Circular Hough Transform does not require a model or feature based approach to circle fitting.

CHAPTER 3

DATASET AND DEVICE

This chapter briefly discusses pupil datasets in Section 3.1, which are publicly available, and we describe their limitations in the context of examining Relative Afferent Pupillary Defect (RAPD). We also discuss the introduced dataset in the same section, which we use for the work presented in this thesis. A device was necessary for consistent acquisition of the data collected for the introduced dataset which includes medical ground truth data necessary for the examination of Relative Afferent Pupillary Defect (RAPD). The developed device is briefly described in Section 3.2.

3.1 Pupil Datasets

Publicly available pupil datasets are summarized in Table 2. The public datasets are primarily utilized for development of gaze-tracking applications. Therefore, they do not include relevant medical data for examination of Relative Afferent Pupillary Defect (RAPD), such as RAPD severity score. The introduced dataset includes videos of pupillary light reflex responses with corresponding medical ground truth labels provided by eye care specialists.

Table 2: Pupil datasets and their attributes.

Dataset	Resolution (pixels)	Environment	Media Type	Medical Ground Truth
Swirski	640x480	Indoor	Image	No
ExCuSe	384x288	Driving, Supermarket	Image	No
EISe	640x480	Indoor, Driving	Image	No
LPW	640x480	Indoor, Outdoor	Video	No
Introduced	1920x1080	Controlled Headset	Video	Yes

3.2 Device: Lab-on-a-headset

A configurable, flexible and portable ocular monitoring device was developed in the OLIVES lab over the duration of one and a half semesters. The device was pertinent to the data collection which is required for the completion of this thesis. Device consists of a hardware and software component, which is briefly described in the follow subsections.

3.2.1 Hardware Component

To ensure consistent data collection, we design the device hardware using an enclosed headset to eliminate or minimize environmental factors such as natural lighting. Early development and current version of the prototypes are shown in Fig. 12. The device is capable of synchronized acquisition of pupillary light reflex responses.

(a) **(b)**

Figure 12: (a) Development device prototype. (b) Prototype utilized in data collection.

3.2.2 Software Component

The device features a graphical user interface (GUI), which is hosted on a local web server by the device. The web server is accessible from any device with Wi-Fi capabiilty, using a browser and web server address. The GUI allows for data-entry by the user to include relevant medical information of the tested subjects. Stimulating lights are configurable using a light sequence editing tool accessible by the device GUI.

CHAPTER 4

ALGORITHMIC EXAMINATION OF RAPD

In this chapter, we describe the algorithmic process that is used to automatically examine Relative Afferent Pupillary Defect (RAPD) from pupillary light reflex response videos. First, in Section 4.1 we briefly describe the examination of RAPD in clinical practice, and then describe an algorithmically equivalent examination of RAPD, derived from the clinical practice. In Section 4.2 we describe the pre-processing steps performed for pupil image enhancement. In Section 4.3 we discuss the process of obtaining pupil size measurements from the pre-processed video data. In Section 4.4 we describe the post-processing steps performed on the collected pupil size measurements, for pupil size measurement noise reduction. In Section 4.5 we explain the method of determining RAPD severity level, which is determined from the pupil size measurements. Finally, also in Section 4.5, we describe evaluation of the algorithm performance, based on the ground truth data available from the introduced dataset discussed in the previous chapter.

The proposed pipeline for algorithmic assessment of RAPD is summarized in Fig. 13. Inputs include pupillary light reflex videos of left and right pupils, and corresponding medical ground truths for RAPD severity scores. Outputs include algorithmically assigned RAPD severity scores, and algorithm performance evaluation score.

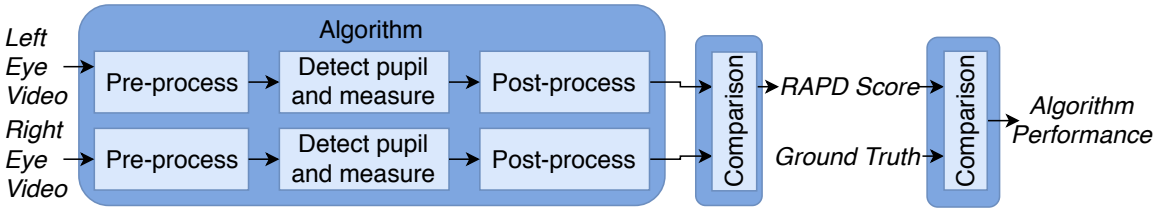


Figure 13: Algorithmic RAPD assessment pipeline.

We divide this chapter into five main sections, namely, algorithmic interpretation of relative afferent pupillary defect, video and image pre-processing, pupil detection and pupil

size measurement, pupil-size-measurement post-processing, and examination of relative afferent pupillary defect.

4.1 Algorithmic Interpretation of Relative Afferent Pupillary Defect

Relative Afferent Pupillary Defect (RAPD) is the condition in which the pupillary light reflex responses of both pupils to a light stimulus differs to one another. This condition is examined in clinical practice by an ophthalmologist or eye care specialist, using a technique known as the swinging flashlight test, described in Section 2.1. These specialists evaluate the severity of RAPD by visually assessing the pupil reactivity to the light stimulus, and the similarity or dissimilarity in the pupillary light reflex responses of both pupils.

Assessing dissimilarity in pupillary light reflex responses requires visual judgement of the physical pupil size differences and the rate of the physical size changes as the pupil reacts to the light stimulus. RAPD severity score is assigned upon assessment, (integers ranging zero to four), which is dependent on the two factors discussed. Score zero corresponds to no RAPD, and score four corresponds to severe RAPD. Clinical procedure pipeline is illustrated in Fig. 14.

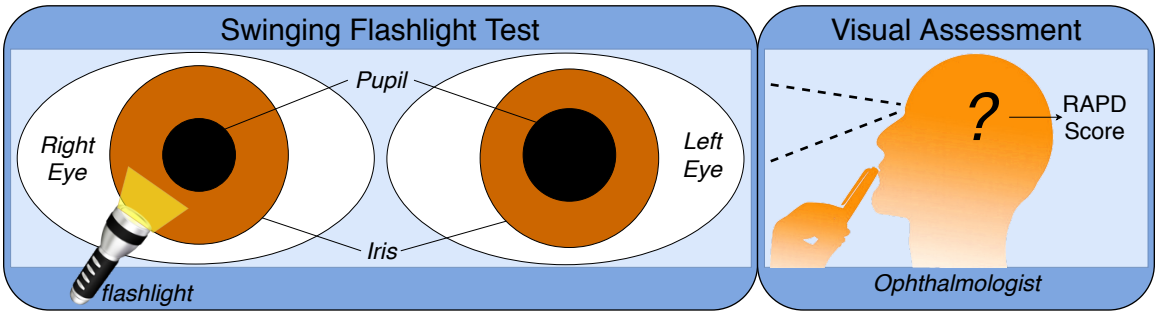


Figure 14: RAPD examination pipeline in clinical practice.

Therefore, from an algorithmic point of view, the pupillary light reflex response can be assessed by measuring the pupil size changes over the duration of the recorded eye videos. Individual image frames of the video correspond to a single pupil size measurement. Measurements are performed in the image domain, therefore, measurement units are in image

pixels. We illustrate the algorithm interpretation of the clinical pipeline, in Fig. 15.

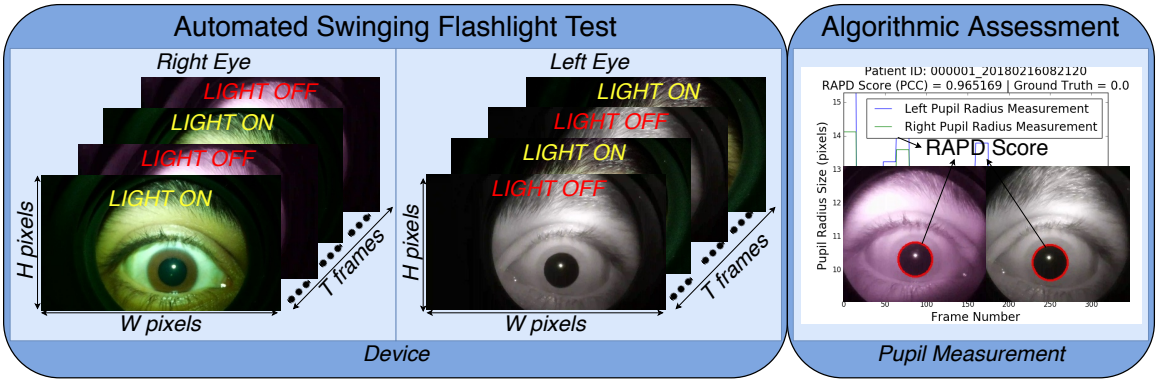


Figure 15: RAPD examination pipeline for algorithmic assessment.

RAPD severity can therefore be quantified by an algorithm, by comparing the similarity of pupil size measurements for both pupils using a similarity measure. A subject without RAPD and without physical eye defects will produce an identical pupillary light reflex response in both pupils. A subject with RAPD scored between one and three inclusive will produce discernible dissimilarity in the pupil responses. However, a subject with severe RAPD (score four) may not produce a pupillary light reflex response, which results in identical pupil size measurements for both pupils. For this case, the lack of pupil reactivity needs to be determined for accurate examination of the RAPD severity level.

The dataset used in this thesis contains light sequence information used in the automated swinging flashlight test. Therefore, we can determine pupil light reactivity based on the timing of the light stimulation. We can use this information in conjunction with the similarity measure to determine the severity level for the special case described here. An example of this case is shown in Fig. 16, in which the stimulated eye shown on the right does not dilate upon stimulation, and the adjacent eye on the left therefore has no pupillary light reflex response.

Subjects may have some other defect or condition, such as a muscle condition in the eye that limits pupil constriction under light stimulation. This may result in a false positive test for RAPD using the algorithmic examination of RAPD, due to differing pupil size

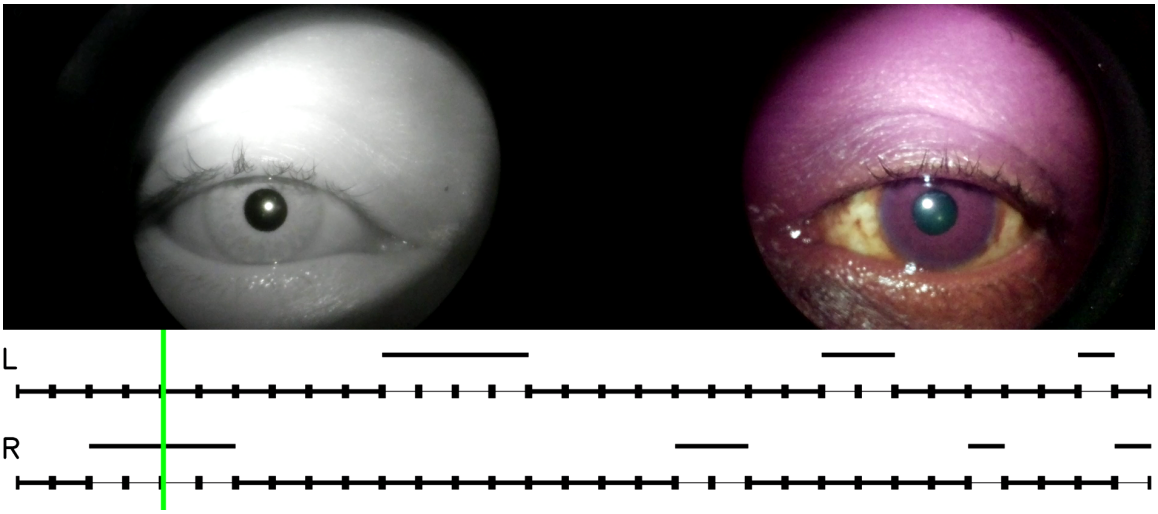


Figure 16: Severe RAPD present in subject with no pupillary light reflex response.

measurement responses. The work presented in this thesis will not account for cases such as this special case. Further investigation is necessary during data collection in order to determine a suitable approach for algorithmic examination of RAPD that is applicable to all possible scenarios of ocular conditions.

The RAPD examination is dependent on the pupil size measurement accuracy. Therefore, in order to effectively obtain the pupil size measurements for an algorithm to examine RAPD adequately, the recorded pupillary light reflex response videos are pre-processed to enhance the pupil characteristics within the frames of video. We describe these techniques in the following section.

4.2 Video and Image Pre-processing

Pupil size measurement from pupillary light reflex response videos requires an understanding of the video characteristics, to effectively detect and measure the pupil. Relevant characteristics include temporal, spatial, and color-space domain properties of the videos. Therefore, the pre-processing steps can be separated into two main subsections, namely, temporal cropping and spatial domain processing. In Fig. 17 we present the proposed pre-processing pipeline.

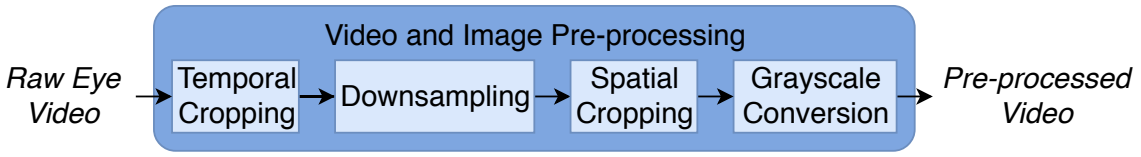


Figure 17: Pre-processing steps performed on input videos.

4.2.1 Temporal Cropping

The pupillary light reflex videos contain two distinct types of image data, namely, color and infrared image data. Color images correspond to sections of the video in which the light stimulus has been turned on. Infrared images correspond to sections of the video in which the light stimulus has been turned off, in which the eyes are illuminated by infrared lighting. We can take advantage of the the pupil's black color or colorless property, and the pupil being the darkest region within the eye. Therefore, the color properties of each video frame can be excluded.

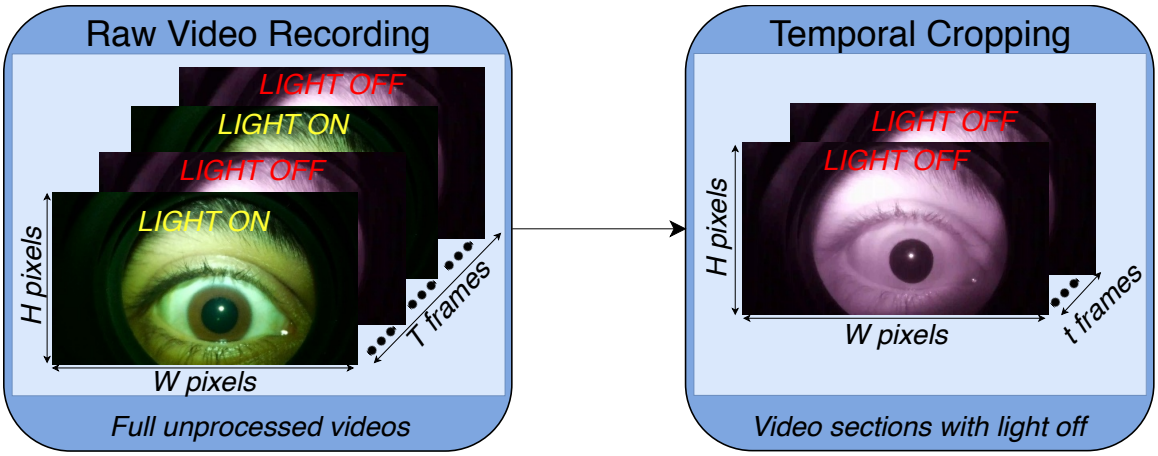


Figure 18: Temporally cropping input videos.

We temporally crop the videos to only include a subset number of frames (t infrared frames) from the total number of frames (T color and infrared frames). We utilize the light sequence information from the introduced dataset to locate the relevant sections of video effectively. In Fig. 18 we illustrate this process of temporal cropping.

4.2.2 Spatial Domain Processing

The dataset videos are captured at full high-definition resolution, 1920 pixles in width by 1080 pixels in height. For the purpose of decreasing processing time, we spatially downsample the videos to 20% of this original resolution. The main purpose in doing so is to reduce computation time, however, downsampling also reduces small noise artifacts from the raw recorded footage.

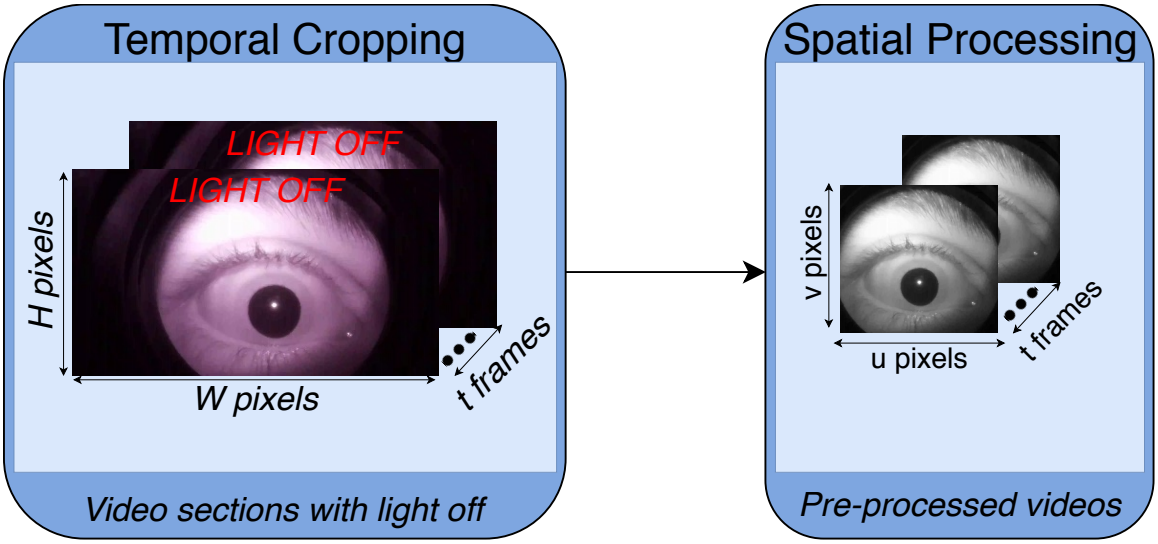


Figure 19: Spatially processing temporally cropped videos.

The device headset used in data collection consistently positions the subject's eyes in a fixed location relative to the cameras. Therefore, the location of the recorded eyes remain constant in the image domain for all videos taken by the same device headset. Therefore, we can consistently localize the pupil region by spatially cropping the frames at predetermined image coordinates specific to the headset. Finally, we convert the resulting frames to grayscale for pupil detection and pupil size measurement. In Fig. 19 we illustrate the spatial domain processing performed on temporally cropped videos.

4.3 Pupil Detection and Pupil Size Measurement

Once the frames have been pre-processed as described in the previous section, we apply the pupil detection and pupil size measurement algorithms. To detect the pupil, we use image properties that are unique from the other regions of the eye. Pupil is the darkest region in the eye, and can be modelled as a circular shape for a 2-dimensional image. Therefore, the pupil can be found by searching for points on the darkest edges of the image frames that correspond to a circle. For this purpose, we incorporate the Circular Hough Transform (CHT), using OpenCV's implementation of CHT algorithm, which is described in [24]. In Fig. 20 we illustrate the pupil detection and pupil size measurement pipeline.

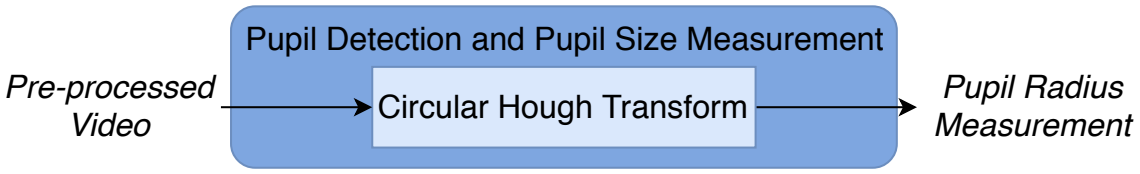


Figure 20: Pupil detection steps performed on preprocessed videos.

4.3.1 Circular Hough Transform

The Circular Hough Transform (CHT) is a variation of the Hough Transform, used to find circles in images, rather than lines. The circles are found by transforming points in the 2-dimensional image plane into right circular cones in the Hough space. Image points that belong to the same circle will have cones in the Hough space that intersect at the same point, and these points correspond to the circle parameters.

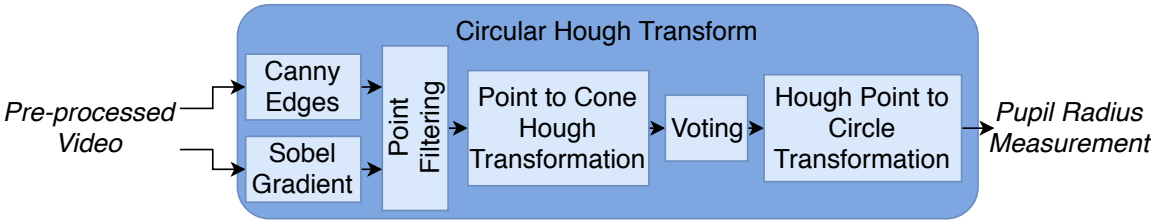


Figure 21: Steps performed in circular hough transform.

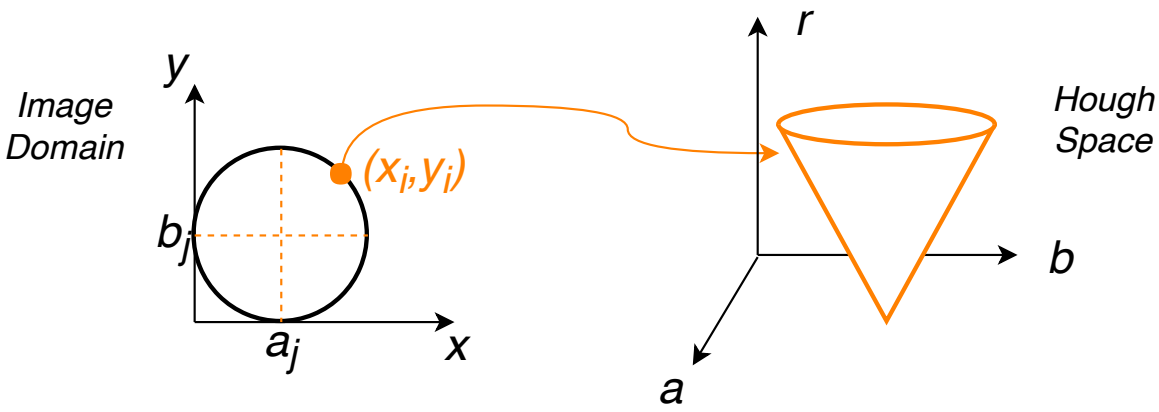


Figure 22: Mapping a point in the image domain to the hough space.

In Fig. 21 we illustrate the various processes within the Circular Hough Transform (CHT) algorithm. Processes include Canny edge detection, Sobel operation, image point constraining, Hough transformation, and voting for circle fitting. Therefore, we apply CHT algorithm for pupil size measurement by modelling the pupil as a circular shape. In Fig. 22 we illustrate the Circular Hough Transformation, in which a point corresponds to a right circular cone.

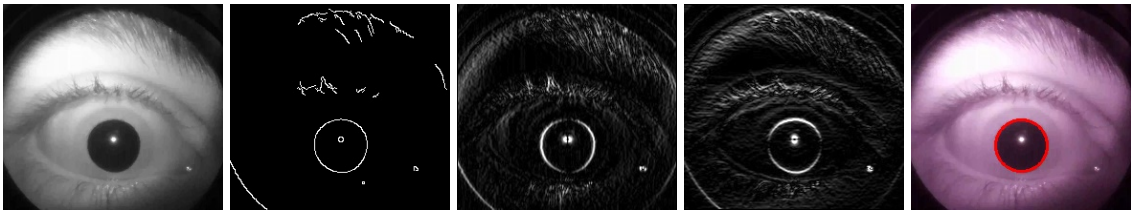
Pupils in 2-dimensional images can be modelled as a circular shape, described by

$$(x - a)^2 + (y - b)^2 = r^2, \tag{2}$$

where (a, b) are the center coordinates of a circle with radius r, and the points (x, y) refer to the edge points within the image that belong to the circle. The Circular Hough Transform (CHT) determines circles within an image by comparing the transformed edge points in the Hough space. An edge point (x_i, y_i) corresponds to a right circular cone in the Hough space, with parameters (a_j, b_j, r_j). As described by [24], the center of the circle must be normal to the points that lie on the circle. Therefore, we can constrain the number of points that are transformed into the Hough space using this criteria.

We illustrate an example of the Circular Hough Transform (CHT) process in Fig. 23. CHT requires edge points for the Hough transformation, which are determined using Canny

edge detection. We expect the pupil to be the darkest region in the eye region, therefore, we use a high threshold gradient for the Canny edge detector. The steepest gradient magnitude in an 8-bit image occurs when moving from the darkest pixel (zero intensity value) to the brightest pixel (255 intensity value), which results in a gradient magnitude of 255. Therefore, we use a Canny gradient threshold ranging between a minimum of 125 pixels and a maximum of 250 pixels, to determine the pupil's edge points. An example result of this operation is presented in Fig. 23b.



(a) **(b)** **(c)** **(d)** **(e)**

Figure 23: (a) The grayscale image after pre-processing. (b) The detected edges from (a) using Canny with gradient threshold range (125,250). (c) Resulting image after performing Sobel operation on (a), in the x-axis direction. (d) Resulting image after performing Sobel operation on (a), in the y-axis direction. (e) The detected pupil after voting in the Circular Hough Transform.

We obtain the gradient information of the eye frames using a Sobel filter. We use a 3×3 Sobel filter in both the x-axis and y-axis directions to calculate the gradients in all directions for each edge point detected. The pupils cover a small percentage of the entire frames, therefore, a larger filter size would lose local gradient information, such as gradients in and around the pupil.

$$G_x = \begin{bmatrix} 2 & 0 & -2 \\ 1 & 0 & -1 \\ 2 & 0 & -2 \end{bmatrix} \qquad G_y = \begin{bmatrix} 2 & 1 & 2 \\ 0 & 0 & 0 \\ -2 & -1 & -2 \end{bmatrix}$$

(a) **(b)**

Figure 24: (a) The 3×3 Sobel filter used to find gradients along the horizontal x-axis direction. (b) The 3×3 Sobel filter used to find gradients along the vertical y-axis direction.

Sobel filters are presented in Fig. 24, for which the resulting images upon applying these filters are presented in Fig. 23c and Fig. 23d respectively. Edge points with non-zero

gradients in all directions are selected for circle detection by the Circular Hough Transform (CHT) algorithm. This approach is implemented as described in OpenCV's implementation of CHT. We utilize the radius parameter of the detected circle as the pupil size measurement.

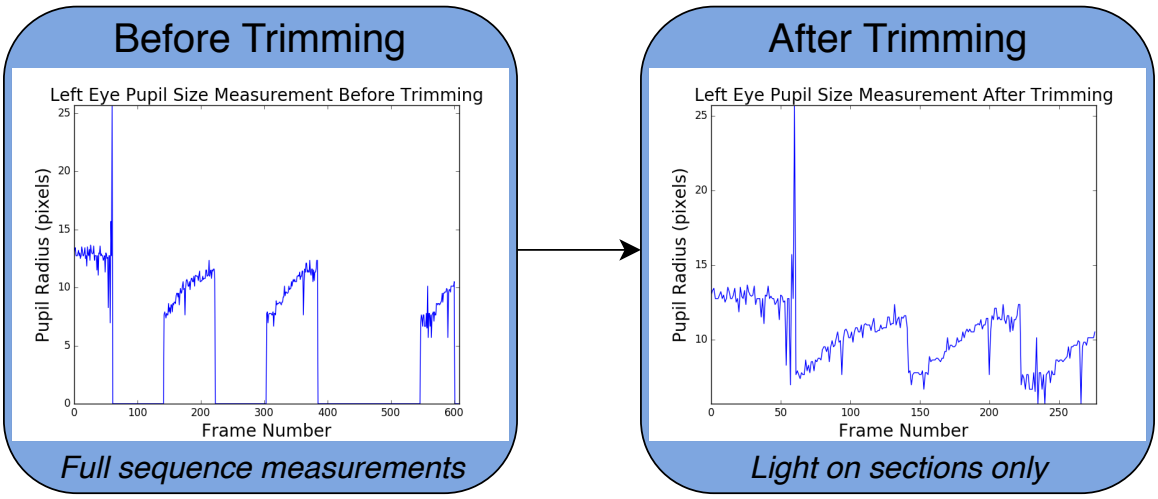


Figure 25: A sample raw pupil radius measurement, emphasizing temporal crop.

CHT is repeated for all pre-processed frames until a complete pupil size measurement curve is obtained, such as that shown in Fig. 25. The plot on the left represents pupil measurements in which the blank sections containing no measurements correspond to color frames of the video which we exclude. The plot on the right represents the collapsed or trimmed pupil size measurements, which consist of only infrared sections of video. We post-process the trimmed measurements before utilizing the measurements in assessment of RAPD.

4.4 Pupil-size-measurement Post-processing

Pupil size measurements can include noise, observed as small and large spikes in the measurement curves. Noise can occur due to false pupil detection on the iris, which shares similar properties of the pupil. In Fig. 25 we observe instances of local noise and large measurement spikes, indicating the detection errors. Therefore, we perform post-processing

techniques on the measured data to eliminate incorrect detection (large spike measurements), and to smooth the measurement curves (local noise). Post-processing techniques include median filtering, thresholding, and moving average. The post-processing pipeline is illustrated in Fig. 26.

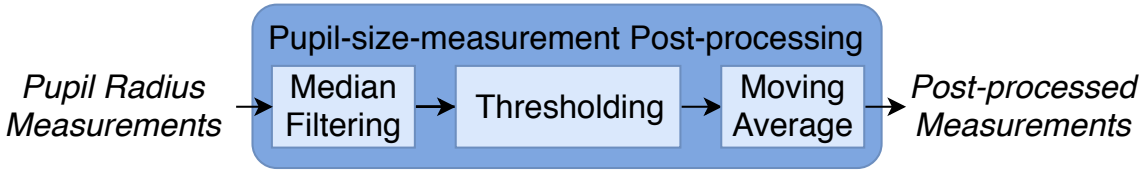


Figure 26: Post-processing steps performed on pupil size measurements.

The overall effect of post-processing is illustrated in Fig. 27, in which we obtain a simplified curve for RAPD assessment. The plot on the left corresponds to the initial trimmed measurements, while the plot on the right corresponds to the post-processed measurements.

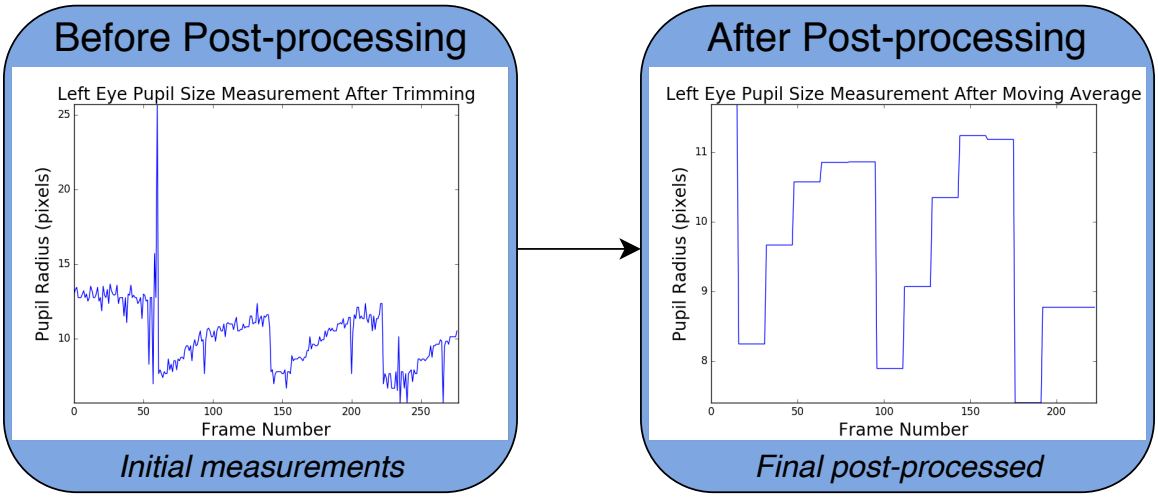


Figure 27: Sample pupil size measurement before and after post-processing.

4.4.1 Median Filtering

The measured pupil sizes contain noise due to imperfect pupil detection, which affects the similarity measure for RAPD assessment. The measurements are in the form of $t \times 1$ vectors, which contain t pupil size measurements for t frames of the pre-processed pupillary

light reflex response videos. The similarity measure for RAPD assessment compares corresponding measurements for left and right pupils. Therefore, local noise in small regions of the measurement can significantly affect the similarity measure, and may result in false positive tests for RAPD. Therefore, we utilize a 1-dimensional median filter to smooth the local noise in the measurement data, while preserving the overall shape and trend of the curves.

A sliding window median operation is applied to the measurement data. The larger the window size, the smoother the curves become. However, a large window size may not preserve the overall trend and shape of the curve. Therefore, we select a window size of 3 measurements, validated by a parameter tuning sweep, presented in Section 5.3. The median operation selects the median measurement within the sliding window region, and replaces the neighborhood measurements with the median measurement.

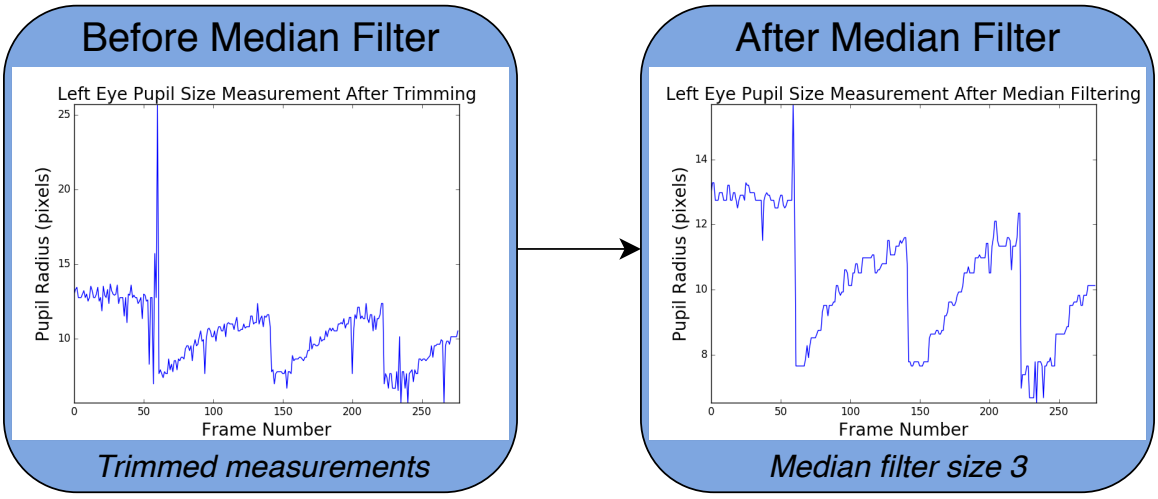


Figure 28: Median filter applied to trimmed measurements.

In Fig. 28, we illustrate the effect of median filtering on the original measurement data. We observe a limitation of median filtering in this figure, in which large measurements spikes are still present. Median filter relies on neighboring data points to identify the median measurement of the localized window region. Therefore, continuous measurement spikes within the sliding window region will not be eliminated by the median operation. Therefore, we require additional post-processing measures to handle these instances.

4.4.2 Thresholding

Thresholding is utilized to filter large continuous measurement spikes which are not filtered by the median filtering stage. We determine the threshold based on the largest measured pupil size. We utilize the initial conditions of the video data to determine the largest pupil size. Each video contains a short buffer period before the automated swinging flashlight test is performed. This ensures maximum pupil dilation before testing. Therefore, we utilize the first 24 frames of pupil measurement data to obtain the largest pupil size. This is validated by a parameter tuning sweep, presented in Section 5.3.

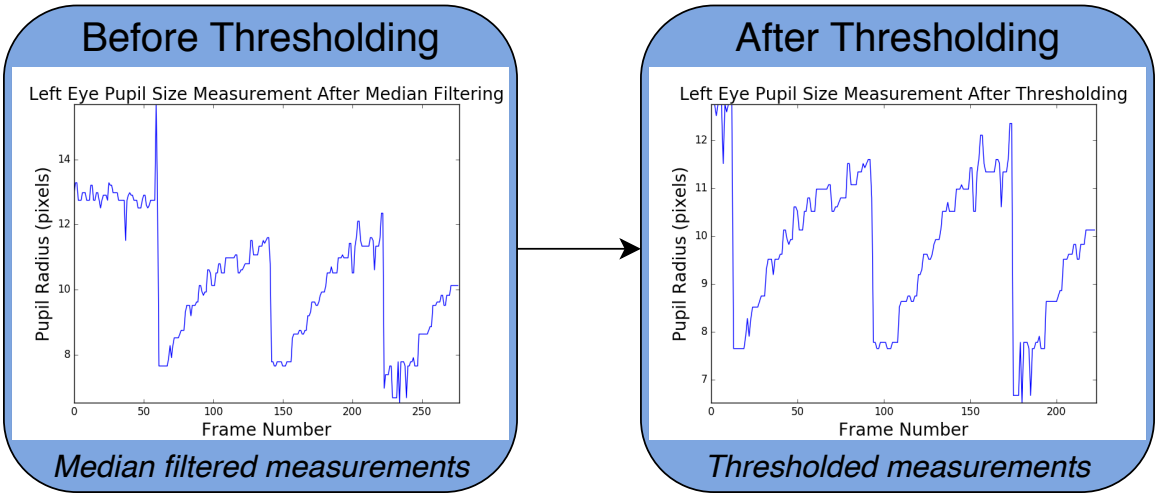


Figure 29: Threshold applied on median filtered measurements.

In Fig. 29 we illustrate the effect of thresholding. We observe the omission of measurements containing large measurement spikes, which is dependent on the average pupil size within the first few frames of the measurement data. However, we also observe local regions containing spikes that are smaller than the smallest pupil size. This may be due to corneal reflection within the pupil, due to the light stimulus. Therefore, further post-processing is utilized to simplify the measurement data even further.

4.4.3 Moving Average

Moving average is utilized to simplify the result of median filtering after tresholding. Moving average simplifies the measurements by utilizing the mean measurement within local regions of the measurement data. For example, moving average with window size 2 applied onto measurement data containing N measurements will result in $N/2$ regions of measurement data that consist of the same measurement value. Therefore, moving average effectively reduces the number of points considered in the similarity measure for RAPD assessment.

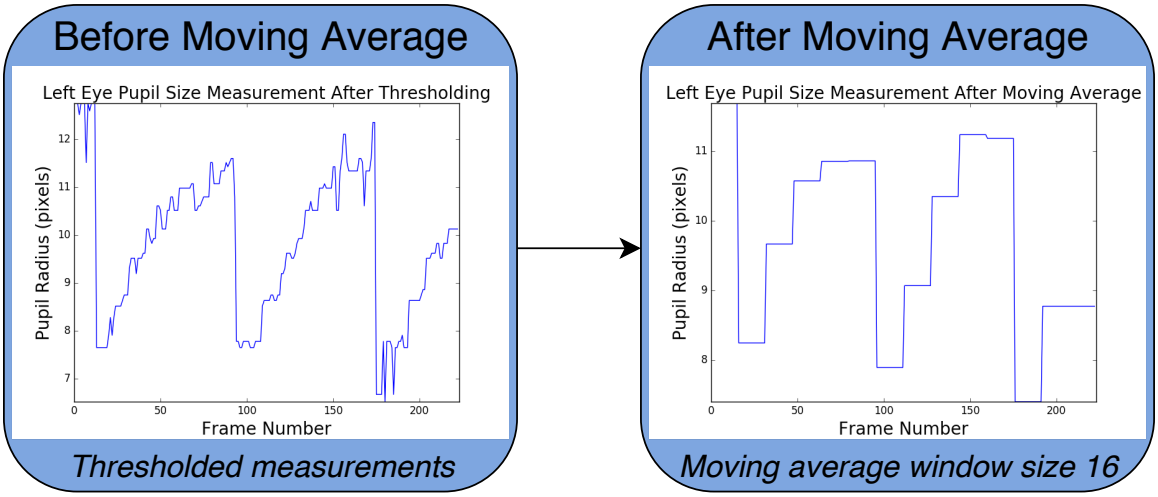


Figure 30: Moving average performed on thresholded measurements.

We illustrate the effect of moving average in Fig. 30. We select a moving average window size of 16 measurements, resulting in $N/16$ regions of measurement data. We validate the selection of this window size by a parameter tuning sweep, presented in Section 5.3. In the figure presented, we observe the elimination of spikes, while preserving the overall trend and shape of the measurement curves. The post-processed measurement data are utilized in the assessment of RAPD.

4.5 Examination of Relative Afferent Pupillary Defect

We determine the severity of Relative Afferent Pupillary Defect (RAPD) by performing a similarity measure on the post-processed pupil size measurement curves. A healthy eye with the absence of RAPD or other eye defects is expected to have identical pupillary light reflex responses in both pupils. An unhealthy eye containing RAPD presents dissimilar pupillary light reflex responses in both pupils. Specifically, dissimilar pupil responses correspond to differing amounts of pupil dilation and constriction, as well as differing rates in pupil size changes.

Pupils that dilate or constrict by the same amount may still present RAPD, which is observable by the differing rates in pupil size changes. For example, the left pupil may dilate increasingly, while the right pupil may dilate decreasingly, indicating a delay in pupillary light reflex response. Therefore, we expect linear correlations in pupil responses to correspond to the absence of RAPD, and non-linear correlations and absence of correlations to correspond to the presence of RAPD. Therefore, we utilize Pearon's Correlation Coefficient for the algorithmic assessment of RAPD, which is a measure of linear correlation only.

Pearson's Correlation Coefficient (PCC) produces a floating point score that ranges between zero and one. PCC score zero corresponds to severe RAPD, in which there is no linear correlation between the pupil size measurements of both pupils. PCC score one corresponds to the absence of RAPD, in which the pupil size measurements of both pupils are perfectly and linearly correlated. The obtained PCC algorithmic RAPD assessment scores for all subjects are validated by comparison with the corresponding medical ground truth scores available in the introduced dataset.

The medical ground truth scores for RAPD range in integers between zero and four. RAPD severity score zero corresponds to the absence of RAPD. RAPD severity score four corresponds to severe RAPD. Therefore, we expect a correlation between the algorithmic

assessment and the medical ground truth, which may or may not be linear. Therefore, we utilize Spearman's Rank Correlation Coefficient (SRCC), which is a measure of monotonically increasing correlation between two signals.

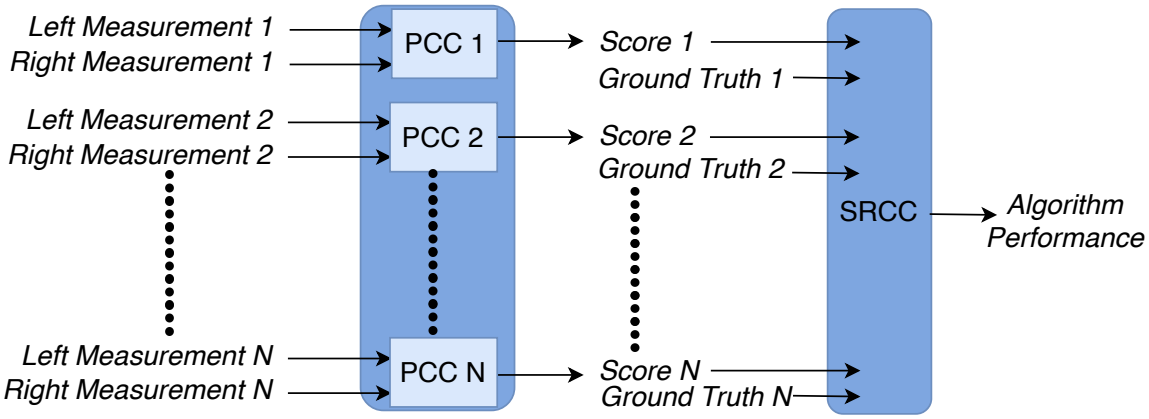


Figure 31: Algorithmic RAPD score by PCC and algorithm assessment by SRCC.

In Fig. 31 we illustrate the process of obtaining RAPD scores by algorithmic assessment. In the same figure we illustrate the evaluation of algorithm performance. Pupil size measurement pairs of N subjects are assessed by performing Pearson's Correlation Coefficient (PCC). The measurement pairs of each subject are mapped algorithmic RAPD scores. The corresponding N medical ground truth RAPD scores are utilized in the evaluation of the algorithm performance, using Spearman's Rank Correlation Coefficient (SRCC). Therefore, we map $N \times 2$ pupil measurements to N algorithmic RAPD scores, and we map $N \times 2$ RAPD scores to a single value, which represents the algorithm performance. PCC and SRCC are briefly described in the follow subsections.

4.5.1 Pearson Correlation Coefficient

The Pearson Correlation Coefficient (PCC) is a measure of linearity between two inputs signals, and in the context of this thesis, these signals refer to the pupil size measurements. This is formulated as

$$PCC = \frac{\sum_{i=1}^{t}(left_i - \overline{left})(right_i - \overline{right})}{\sqrt{(\sum_{i=1}^{t}(left_i - \overline{left})^2)}\sqrt{(\sum_{i=1}^{t}(right_i - \overline{right})^2)}} \tag{3}$$

in which $left$ and $right$ refer to the pupil size measurements for the left and right pupils respectively. PCC comparison is performed on all t post-processed measurements for each subject. The measurements are obtained from t pre-processed video frames.

4.5.2 Spearman Rank Correlation Coefficient

The Spearman's Rank Correlation Coefficient (SRCC) is a measure of the monotonic relationship between two scores or rank inputs. This is formulated as

$$SRCC = 1 - \frac{6 \sum_{i=1}^{N} (Algorithm_i - Manual_i)^2}{N \cdot (N^2 - 1)} \tag{4}$$

in which $Algorithm$ and $Manual$ refer to the RAPD scores obtained from the algorithm and the medical ground truths respectively. N refers to the number of subjects tested.

CHAPTER 5

RESULTS AND DISCUSSION

In this chapter we present the algorithm performance for automated RAPD examination. We first discuss the effect of the proposed pipeline on the the algorithm performance, and identify the strongest contribution to the performance improvement. We then present the trend in performance improvement by analyzing the various combinations of the pipeline stages, and show the scatter plot of the best result. We discuss the cases in which the algorithm fails to detect the pupil correctly, by analyzing case examples that deviate from the trend seen in the best performance scatter plot. Finally, we briefly discuss parameter tuning and selection for the post-processing stage of the algorithm pipeline.

5.1 Algorithm Performance Trend in Relation to Proposed Pipeline

In this section we discuss the overall effect of the proposed pipeline on the algorithm performance, as well as the individual contribution of the various pipeline stages. In Table 3, we present the effect of the proposed pipeline on the algorithm performance, by indicating the algorithm performance after adding each stage of the proposed pipeline sequentially. It is evident from these results that there is an increasing trend in performance improvement

Table 3: Sequential algorithm block contribution to algorithm performance.

Temporal Crop	Downsample	Spatial Crop	Median Filter	Threshold	Moving Average	Algorithm Performance
✗	✗	✗	✗	✗	✗	0.362455
✓	✗	✗	✗	✗	✗	0.064810
✓	✓	✗	✗	✗	✗	0.093614
✓	✓	✓	✗	✗	✗	0.096015
✓	✓	✓	✓	✗	✗	0.206432
✓	✓	✓	✓	✓	✗	0.775319
✓	✓	✓	✓	✓	✓	0.842529

with the sequential addition of each stage in the proposed pipeline. We also notice a significant performance improvement with the addition of the threshold stage, increasing the algorithm performance from 21% to 78%.

Table 4: Individual algorithm block contribution to algorithm performance (inclusion).

Temporal Crop	Downsample	Spatial Crop	Median Filter	Threshold	Moving Average	Algorithm Performance
✗	✗	✗	✗	✗	✗	0.362455
✓	✗	✗	✗	✗	✗	0.064810
✗	✓	✗	✗	✗	✗	0.412863
✗	✗	✓	✗	✗	✗	0.316848
✗	✗	✗	✓	✗	✗	0.194430
✗	✗	✗	✗	✓	✗	0.717710
✗	✗	✗	✗	✗	✓	0.237636

In Table 4, we observe that the inclusion of only the threshold stage results in the most significant performance improvement on the baseline performance, from 36% to 72%. In Table 5, we observe that by excluding only the threshold stage from the complete algorithm pipeline, the algorithm performance drops significantly, from 84% to 15%. Therefore, the threshold stage is observed as the most significant component of the algorithm pipeline, in terms of improving algorithm performance. However, the overall performance improvement trend is contradicted by the performance drop upon the addition of temporal crop.

Table 5: Individual algorithm block contribution to algorithm performance (exclusion).

Temporal Crop	Downsample	Spatial Crop	Median Filter	Threshold	Moving Average	Algorithm Performance
✓	✓	✓	✓	✓	✓	0.842529
✗	✓	✓	✓	✓	✓	0.549684
✓	✗	✓	✓	✓	✓	N/A
✓	✓	✗	✓	✓	✓	0.837728
✓	✓	✓	✗	✓	✓	0.703308
✓	✓	✓	✓	✗	✓	0.146422
✓	✓	✓	✓	✓	✗	0.775319

We observe in Table 3 that the initial inclusion of temporal crop results in a performance drop from 36% to 6%. The drop in performance due to the temporal crop stage

can be explained by analyzing the measurement data before and after temporal cropping, and relating those measurements to the corresponding video data. Before temporal cropping, the algorithm produces noisy measurement data such as that shown in Fig. 32. The noise seen in the measurement data is due to inconsistent and inaccurate pupil detection in the sections of videos in which the light is turned on. However, this would suggest that temporal crop applied to these sections should improve the algorithm performance.

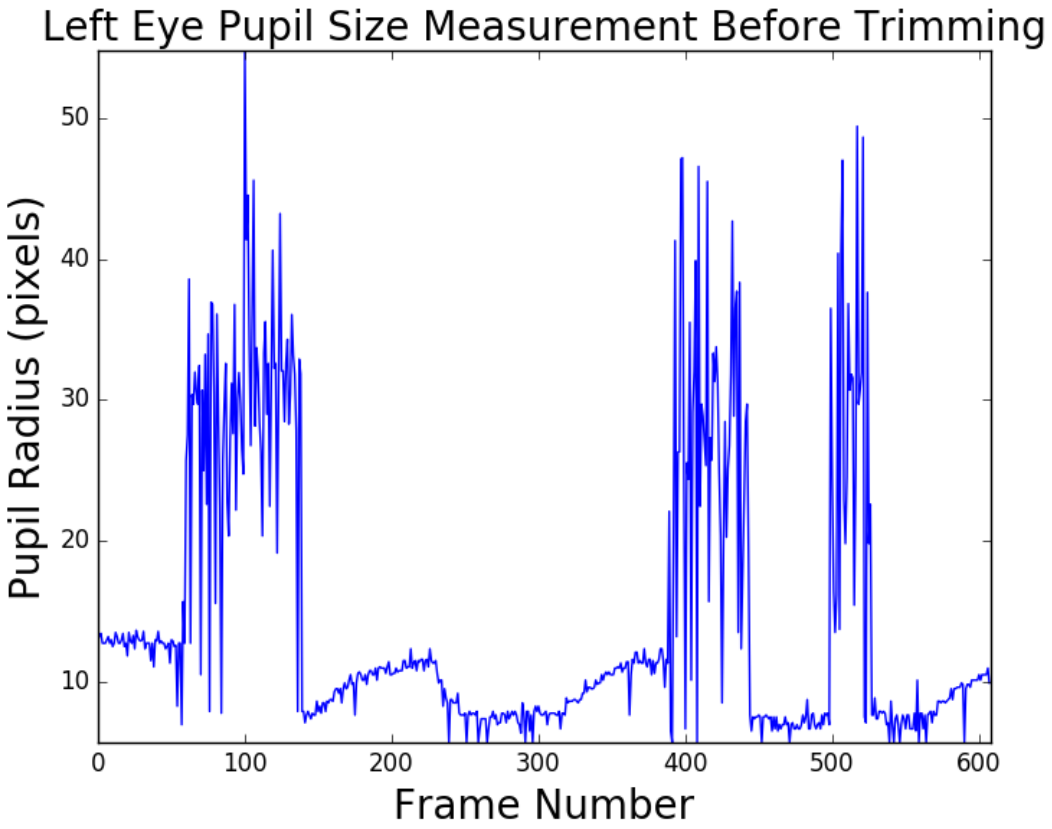


Figure 32: Pupil size measurement without temporal cropping.

However, we note that there are significant portions of noisy measurements, and these measurements are consistently random such that there is some correlation between the noise components in both left and right pupil measurements. This explains the initial baseline performance of 36%. After temporally cropping out the sections of video that cause this noise, we obtain measurement data such as that shown in Fig. 25 in the previous chapter.

Table 6: Algorithm performance pipeline sweep.

Temporal Crop	Downsample	Spatial Crop	Median Filter	Threshold	Moving Average	Algorithm Performance
✓	✗	✓	✗	✗	✗	0.021603
✓	✗	✗	✓	✗	✓	0.060009
✓	✗	✓	✗	✗	✓	0.062410
✓	✗	✗	✗	✗	✗	0.064810
✓	✗	✗	✗	✗	✓	0.067210
✓	✗	✗	✓	✗	✗	0.081612
✓	✗	✓	✓	✗	✗	0.088814
✓	✓	✗	✗	✗	✗	0.093614
✓	✓	✗	✓	✗	✓	0.093614
✓	✗	✓	✓	✗	✓	0.096015
✓	✓	✓	✗	✗	✗	0.096015
✓	✓	✗	✗	✗	✓	0.105616
✓	✓	✓	✗	✗	✓	0.120018
✓	✓	✓	✓	✗	✓	0.146422
✗	✗	✗	✓	✗	✗	0.194430
✗	✗	✓	✗	✗	✓	0.194430
✓	✓	✓	✓	✗	✗	0.206432
✓	✓	✗	✓	✗	✗	0.208832
✗	✗	✗	✗	✗	✓	0.237636
✗	✗	✓	✓	✗	✗	0.254439
✗	✗	✗	✓	✗	✓	0.264040
✗	✗	✓	✓	✗	✓	0.290444
✗	✗	✓	✗	✗	✗	0.316848
✗	✓	✓	✗	✗	✗	0.355254
✗	✗	✗	✗	✗	✗	0.362455
✗	✓	✗	✗	✗	✗	0.412863
✗	✓	✓	✓	✗	✗	0.412863
✗	✓	✗	✗	✗	✓	0.441668
✗	✓	✗	✓	✗	✗	0.441668
✗	✓	✓	✗	✗	✓	0.441668
✗	✓	✗	✓	✗	✓	0.451269
✗	✓	✓	✓	✗	✓	0.451269
✗	✓	✓	✓	✓	✓	0.549684
✓	✗	✗	✗	✓	✗	0.571287
✓	✗	✓	✗	✓	✗	0.571287
✗	✓	✗	✓	✓	✓	0.595291
✗	✗	✗	✓	✓	✓	0.619295
✗	✗	✓	✓	✓	✓	0.650500
✓	✓	✗	✗	✓	✗	0.662501
✓	✓	✓	✗	✓	✗	0.662501
✓	✗	✗	✓	✓	✗	0.676904
✓	✗	✓	✓	✓	✗	0.681704
✓	✓	✗	✗	✓	✓	0.693706
✓	✓	✓	✗	✓	✓	0.703308
✗	✗	✗	✗	✓	✗	0.717710
✗	✗	✗	✓	✓	✗	0.717710
✗	✗	✓	✗	✓	✗	0.717710
✗	✓	✗	✗	✓	✗	0.722511
✗	✓	✗	✓	✓	✗	0.722511
✗	✓	✓	✗	✓	✗	0.722511
✗	✓	✓	✓	✓	✗	0.722511
✗	✗	✓	✗	✓	✓	0.739313
✗	✗	✓	✓	✓	✗	0.741713
✗	✗	✗	✗	✓	✓	0.748915
✓	✓	✗	✓	✓	✗	0.775319
✓	✓	✓	✓	✓	✗	0.775319
✗	✓	✗	✗	✓	✓	0.787320
✗	✓	✓	✗	✓	✓	0.787320
✓	✓	✗	✓	✓	✓	0.837728
✓	✓	✓	✓	✓	✓	0.842529

Although the inclusion of temporal crop drops the performance to less than 7%, in Table 5, we note the significance of temporal crop on the performance improvement, by excluding temporal crop from the complete algorithm pipeline. We observe a performance drop from 84% to 55%, which suggests that the elimination of the noisy measurements results in an overall improvement to the algorithm performance. This also validates the observation that baseline performance is due to correlation in the noisy components in both left and right pupil measurements.

In Table 6 we present all possible combinations of the pipeline stages that produce correlation scores. The table is sorted by algorithm performance to highlight the performance improvement trend. It is evident from this table that the most significant component of the pipeline is the thresholding stage. We also note the significance of temporal crop on the overall algorithm performance.

5.2 Best Performance Result

In this section we present the best obtained algorithm performance and discuss the cases in which the algorithm fails. In Fig. 33, we present the scatter plot of the best result. Algorithmically produced RAPD examination scores are plotted along the x-axis, and the medical ground truth RAPD scores are plotted along the y-axis. We observe a trend in which higher algorithm RAPD scores (x-axis) are reflected by lower manually assigned RAPD scores (y-axis). However, there are some discrepancies seen in the scatter plot. Specifically, we discuss subjects 17, 18, 34, 8 and 9, that deviate from the overall trend. We analyze the videos of these subjects, highlighting discrepancies observed in both the pupil measurements and the corresponding frames of video for the measurements.

Visual investigation of the videos for subjects 17, 18 and 34, indentifies sources of error including, excessive blinking, corneal reflection, and pupil shape. We analyze discrepancies observed in the measurement data, in conjunction with the corresponding sections of the subject videos, to identify these sources of error. We illustrate this analysis, in Fig.

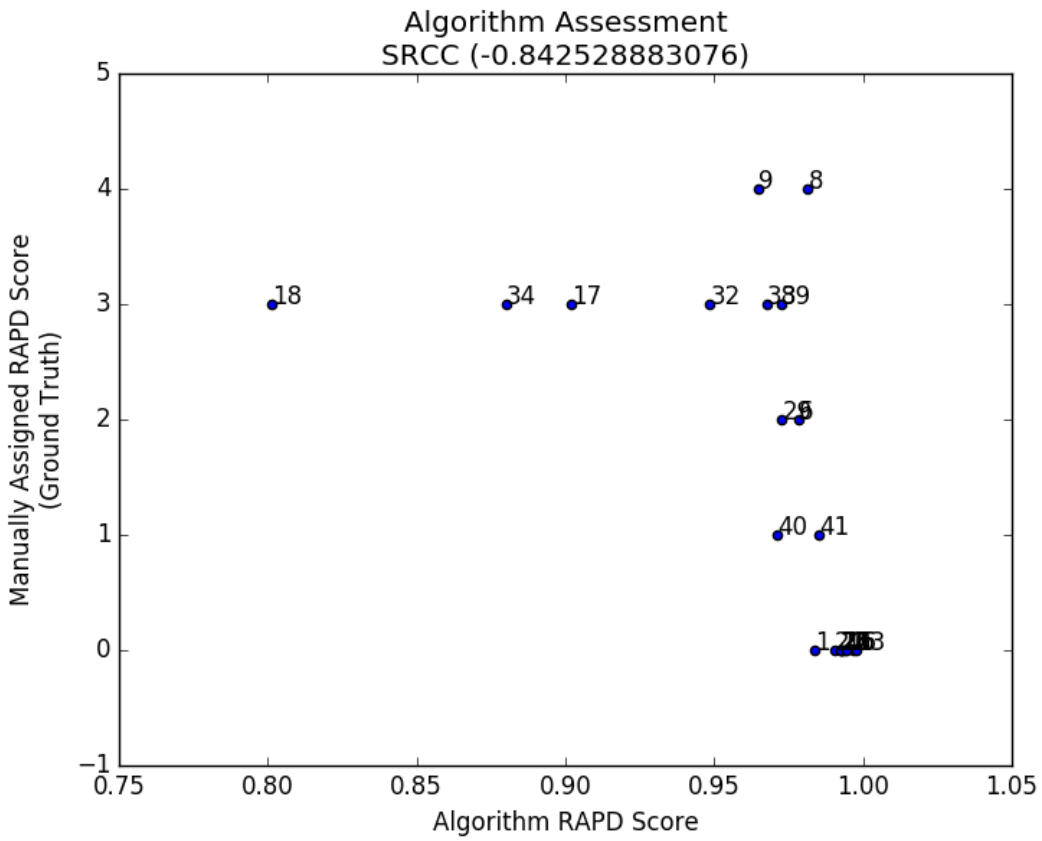


Figure 33: Algorithm performance after selecting best parameters.

34, Fig. 35 and Fig. 36, in which sample frames containing these instances of error are presented with their corresponding measurement discrepencies.

In Fig. 34 we present the analysis for subject 17, in which the post-processed pupil size measurements are compared to the raw measurements before post-processing. The comparison identifies the sections of frames in which errors occur, for which we present selected frames to highlight the measurement discrepancies. We highlight the measurement discrepancies by orange circles, and use arrows to indicate the corresponding sample frames for these discrepancies. Subject 17 presents errors mostly due to corneal reflection overlap with the pupil, which causes either false pupil detection on the reflection, or a failure to detect the pupil altogether. This is due to the non-circular contour formed by the overlap of the corneal reflection on the pupil, illustrated by the Canny edges of the sample frames in the same figure.

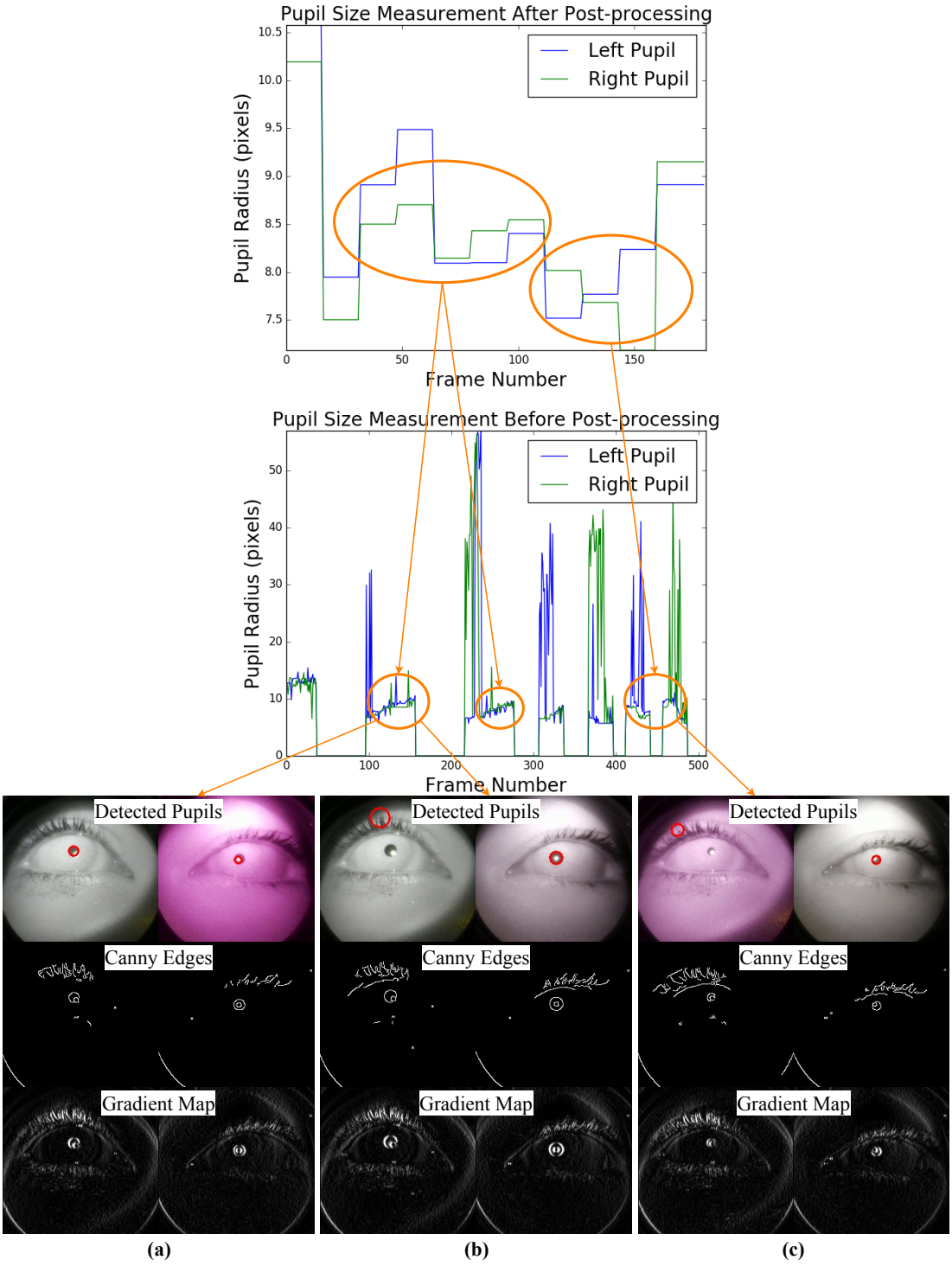


Figure 34: (a) Incorrect pupil detection in left pupil (right image) due to corneal relfection. (b) Incorrect pupil detection due to contour points in eye lashes not filtered out in right eye (left image). (c) Severe corneal reflection overlap on pupil in right eye (left image).

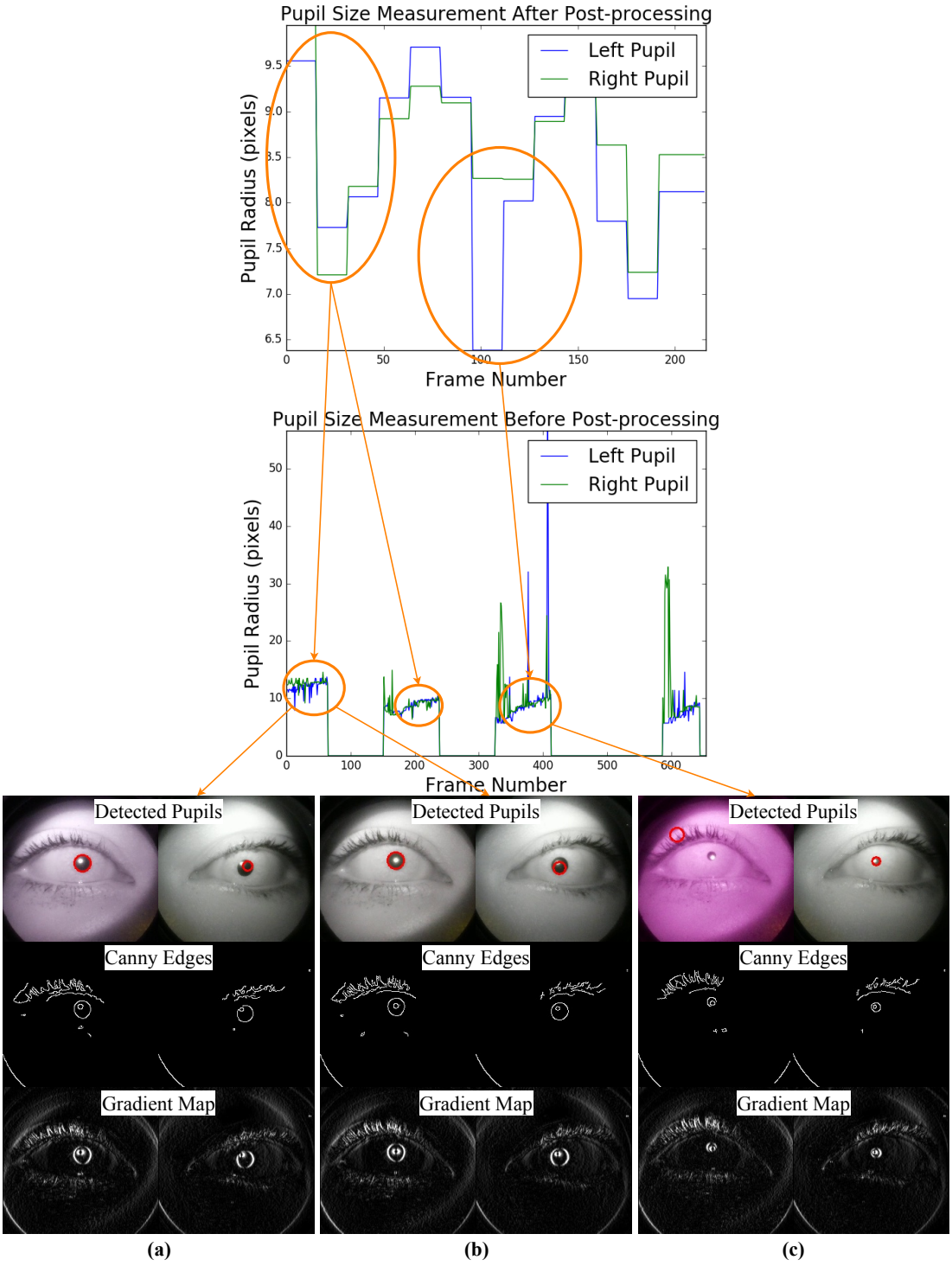


Figure 35: (a) Non-circular shape seen in contour for left pupil (right image) due to corneal reflection overlap. (b) Incorrect pupil detection due to corneal reflection overlap on pupil in left pupil (right image). (c) Severe corneal reflection overlap on pupil in right eye (left image) and availability of eye lash contour points for incorrect pupil detection.

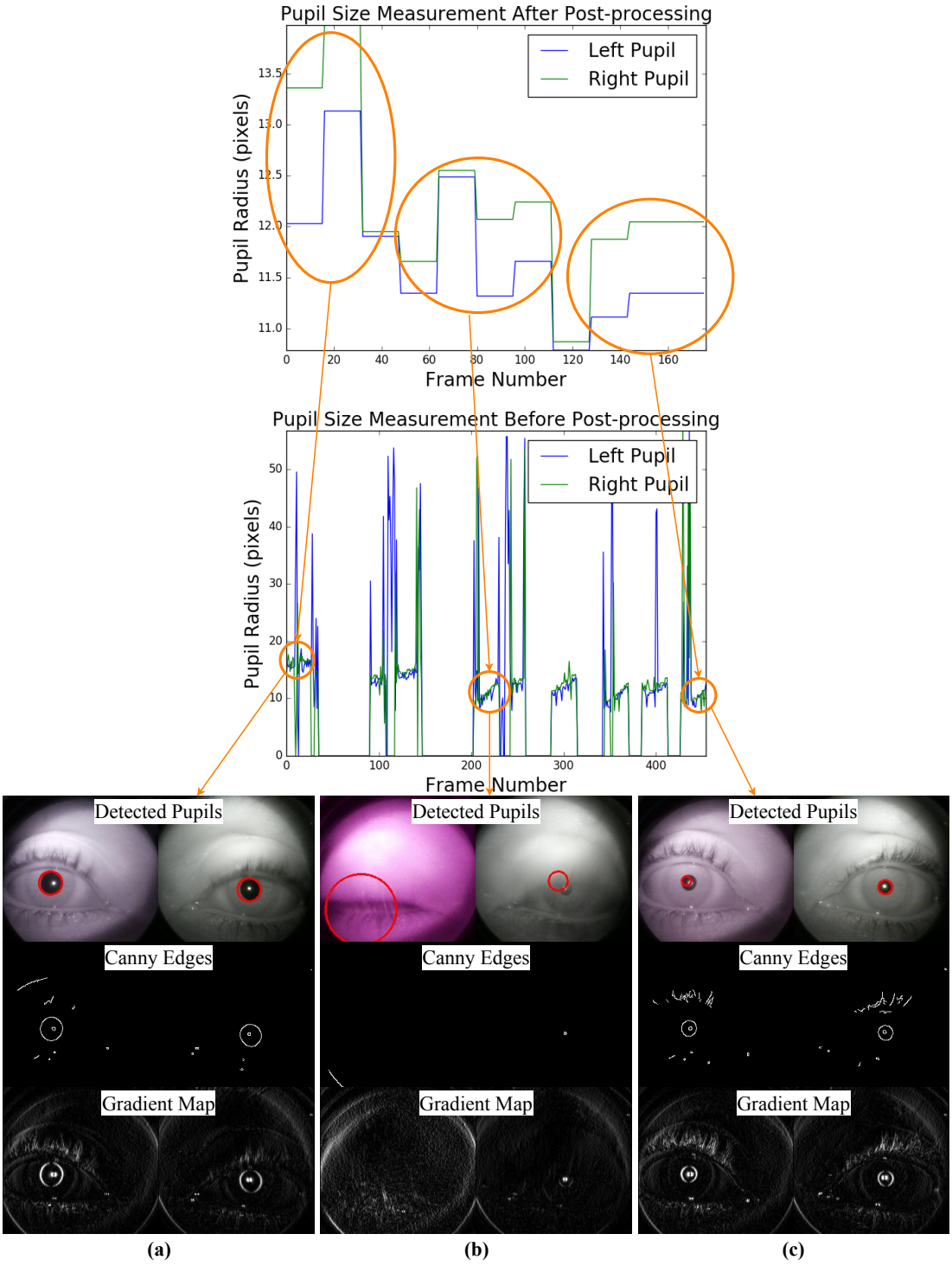
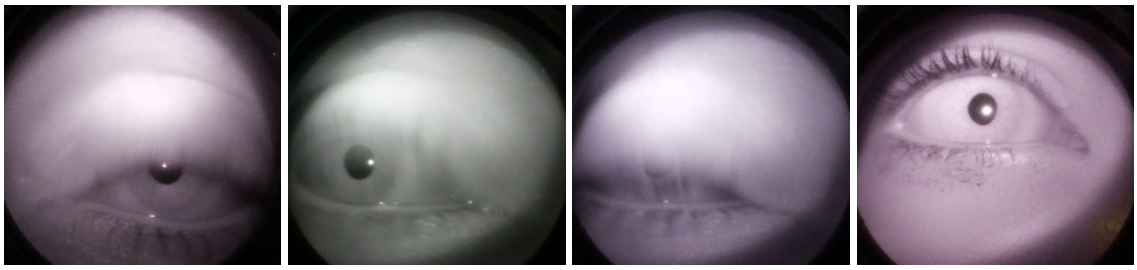


Figure 36: (a) Slight non-circular pupil shape causing spikes in measurement data due to inconsistent circle fitting. (b) Excessive blinking in both eyes (both images) with no visible pupils, but contour points still available due to other regions of image. (c) Inconsistent pupil detection due to corneal reflection in right eye (left image).

In Fig. 35 we present the analysis for subject 18, in which the post-processed pupil size measurements are compared to the raw measurements before post-processing. We present sample frames at the corresponding highlighted areas, indicated by orange circles and arrows, illustrating the sample frames corresponding to the measurement discrepancies. Subject 18 presents similar errors to that of subject 17, in which errors are typically due to corneal reflection overlap with the pupil. Corneal reflection overlap with the pupil causes non-circular edge contours observed in the Canny edge detection result, which affects the result of the Circular Hough Transform for circle detection.

In Fig. 36 we present the analysis for subject 34, in which the post-processed pupil size measurements are compared to the raw measurements before post-processing. We illustrate sample frames which correspond to the highlighted measurement discrepancies, indicated by the orange circles and arrows. Subject 34 presents errors due to excessive blinking, and errors due to a slightly non-circular pupil shape in the subject's left eye. Other sources of error include corneal reflection overlap with the pupil, as observed similarly in the previous subjects discussed.



(a) **(b)** **(c)** **(d)**

Figure 37: (a)–(c) Excessive blinking with motion blur (subject 34). (d) Non-circular pupil shape (subjects 17 and 18).

We present sample frames of challenging conditions for pupil detection in Fig. 37, which include excessive blinking and non-circular pupil. These conditions may be unavoidable, due to the health conditions of the subjects. Therefore, the current capability of the proposed algorithm is not able to account for these cases. This requires a more robust solution to pupil detection and tracking, which is capable of detecting occluded pupils, mo-

tion blurred pupils, and non-circular shaped pupils. However, challenging conditions such as corneal reflection overlap with the pupil can be overcome by pre-processing of corneal reflection detection and removal.

Pupil measurements and corresponding video data for subjects 8 and 9 present no discrepancies, unlike the previous subjects discussed. Therefore, the observed high correlation in these subjects requires a further understanding of RAPD severity score four. We present the post-processed pupil measurement for subject 9 in Fig. 38.

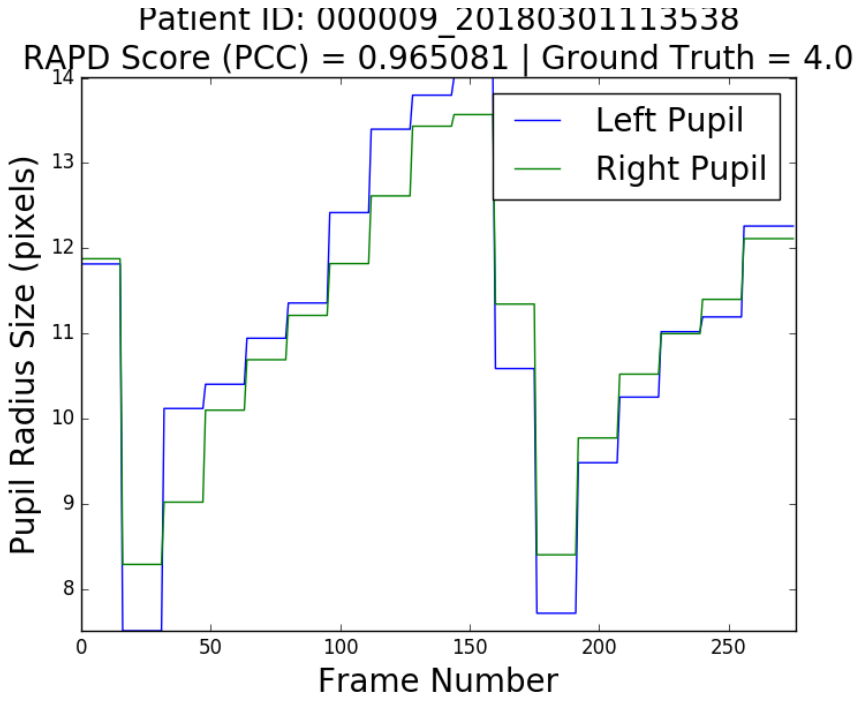


Figure 38: High correlation for pupil response observed in subject 9.

Subjects with RAPD severity score four can include severe cases in which the pupils do not respond to light stimulation. Therefore, pupillary light reflex response may not occur in the adjacent eye, which results in identical pupil measurements for both pupils. Therefore, a similarity measure on the pupil responses by itself is insufficient in examining RAPD. Additional assessment of the pupil's reactivity to the light stimulus is required, in conjunction with the pupil measurement correlation score. Pupil reactivity can be determined by determining pupil constriction upon light stimulation. Therefore, the algorithm produces

a correct assessment of the correlation in measurements for subjects 8 and 9, and the additional information of the pupil's reactivity indicates the corresponding RAPD severity score four.

5.3 Parameter Tuning

In this section we present the parameter tuning sweeps for parameter selection in the post-processing stage of the proposed algorithm pipeline. In Fig. 39 we present the sweep for median filter window size, in which a window size of 3 measurements produces the best performance. This matches the prediction and explanation provided in the previous chapter, in which we explain that a small window size is expected to capture small noise components in the measurement data, while a larger window size over-smooths the signal to the extent at which we lose signal information.

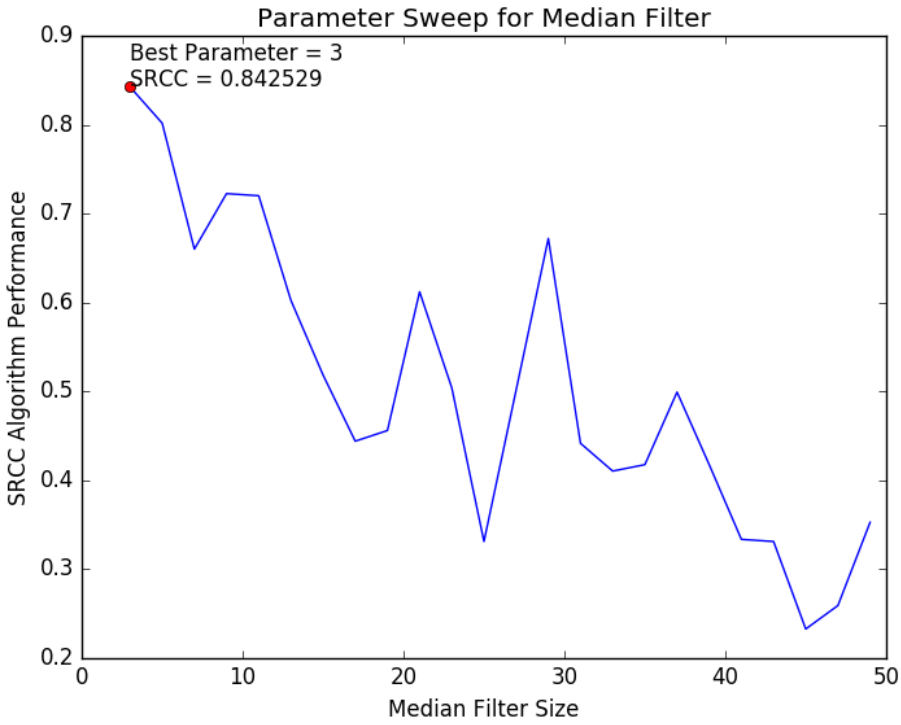


Figure 39: Parameter sweep for median filter size.

In Fig. 40 we present the sweep for thresholding window range, in which window range

of 24 frames produces the best performance. This coincides with the explanation presented in the previous chapter. The purpose of thresholding is to eliminate large spikes in the data due to incorrect pupil detection on the larger iris within the image, as well as other artifacts that lead to large spikes. We expect the largest dilation size of the pupil to occur at the start of the video, before the testing sequences. The range of values used to determine the appropriate maximum threshold is based on this initial section of the videos. Therefore, we expect the range to be large enough to capture the entirety of the initial pupil size measurements before testing, while remaining small enough such that no measurements beyond this initial section are included.

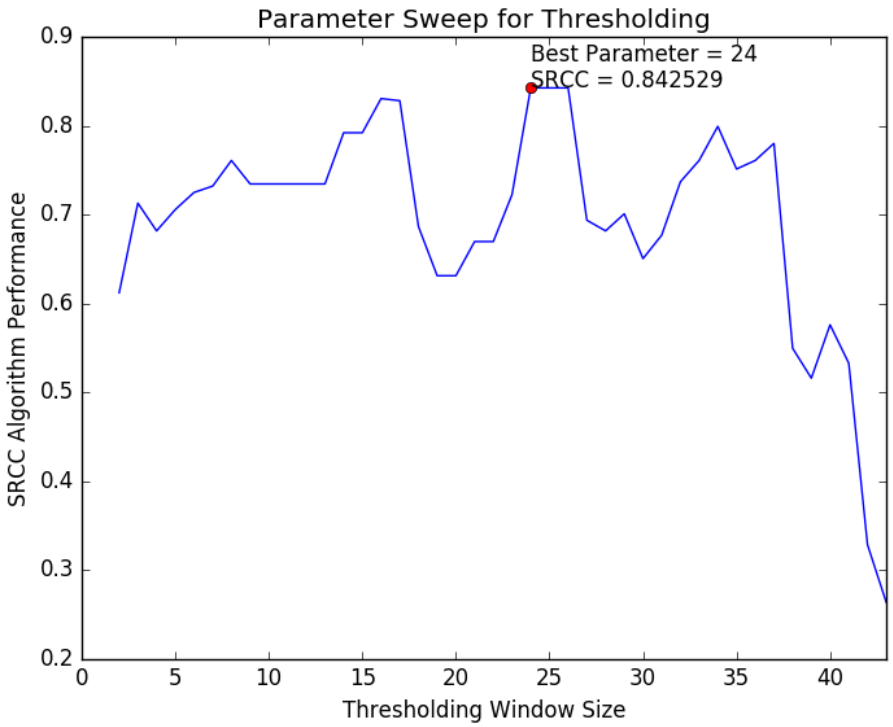


Figure 40: Parameter sweep for thresholding window size.

In Fig. 41 we present the sweep for moving average window size, in which window size of 16 produces the best performance. The purpose of the moving average is to simplify the signal to capture the major pupil dilation or constriction size changes. Therefore, we expect an appropriate window size to be less than the length of a change in light sequence, but also

large enough to average the various sections of the measurement data without losing the overall trend and shape of the measurements curves. This is observed in the window size obtained, in which larger window sizes result in poor performance. Large window size effectively averages the entire measurement data. We recall that moving average reduces the number of data points for the similarity measure, such that a window size w reduces N measurements into N/w regions of average measurements. Therefore, the larger the window size, the lesser number of unique measurements available for comparison.

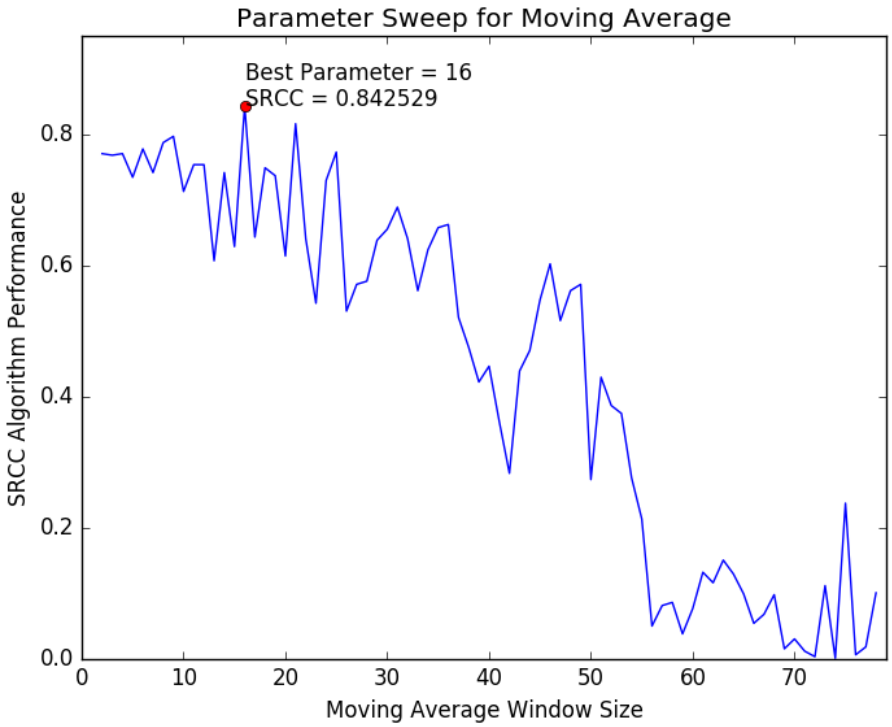


Figure 41: Parameter sweep for moving average window size.

CHAPTER 6
CONCLUSION

This chapter discusses the contributions and future work prospects of this thesis. In Section 6.1 we discuss the contributions of the proposed pipeline for automated examination of Relative Afferent Pupillary Defect (RAPD), and we compare with the state-of-the-art algorithms discussed in the literature. We then discuss in Section 6.2 the future work directions that can be taken from the work presented in this thesis.

6.1 Contributions

The proposed algorithm is a three stage pipeline which ensures accuracy, consistency and automation in examining Relative Afferent Pupillary Defect (RAPD). We propose an objective assessment of RAPD by utilizing a dissimilarity or similarity measure (Pearson's Correlation Coefficient) on the obtained pupil size measurements, in addition to assessing the reactivity of the pupil to a light stimulus. The algorithm for obtaining the pupil size measurements includes pre-processing techniques for pupil image enhancement, pupil detection for pupil contour estimation, and post-processing techniques for pupil measurement noise removal.

We initially investigate the practice of existing methods and their corresponding pre-processing techniques, and identify the common characteristics that are used to perform pupil image enhancement. We also investigate state-of-the-art algorithms for pupil detection and tracking, to highlight common algorithmic techniques that are used to determine the pupil's contour. The proposed pre-processing pipeline includes temporal cropping, spatial cropping and spatial downsampling.

We compare our algorithm with state-of-the-art algorithms for pupil detection in Table 7. The proposed algorithm utilizes temporal characteristics of the video data, unlike

the state-of-the-art-algorithms which rely on static pupil image data. However, we do not utilize other algorithmic techniques such as morphological operations, ellipse fitting, and color space manipulation. The proposed algorithm for pupil detection incorporates the Circular Hough Transform, which utilizes Canny edge detection, Sobel operator, and Hough transformation, to selectively choose edge points in the pupil image for circle detection and fitting.

Table 7: State-of-the-art and the proposed algorithm for this thesis.

Algorithms	Adachi	Masi	Starburst	Swirski	Pupil Labs	SET	ExCuSe	ElSe	Proposed
Binary Threshold	✓	✓	✓	✓	✗	✓	✗	✗	✓
Canny Edges	✗	✗	✗	✓	✓	✗	✓	✓	✓
Morphological Operators	✓	✗	✓	✓	✗	✗	✓	✓	✗
Sobel Operator	✗	✓	✗	✗	✗	✗	✗	✗	✓
Ellipse Fitting	✗	✗	✓	✓	✓	✓	✓	✓	✗
Circle Fitting	✓	✓	✗	✗	✗	✗	✗	✗	✓
Color Space	✓	✗	✗	✗	✗	✗	✗	✗	✗
Temporal Characteristics	✗	✗	✗	✗	✗	✗	✗	✗	✓

We do not utilize color space manipulation techniques due to availability of colorless image data. We choose to eliminate the portions of video data containing color, to simplify the algorithm detection of the pupil. Color portions of the video data present inconsistencies among the subjects, such as differing iris and skin colors, which affects parameter tuning of color enhancement techniques. Therefore, color space manipulation is avoided in the proposed pipeline due to the availability of infrared image data.

Morphological operations require fine parameter tuning, which is dependent on the image properties. For example, pupil sizes can vary significantly between individuals, which affects the amount of morphological open and close operations that may be necessary in pre-processing. This can be impractical in a real world application, therefore, we avoid these techniques in the proposed algorithm.

Ellipse fitting is an alternative approach to pupil contour fitting by circle fitting. As discussed in the literature, ellipse fitting is sensitive to detected pupil feature points, whereas the Circular Hough Transform is more robust to false feature points. Therefore, we utilize the robustness of circle fitting for greater consistency, despite the higher accuracy potential of contour fitting using ellipse fitting.

6.2 Future Work

The proposed algorithm pipeline achieves a performance of 84% accuracy for RAPD examination on the available test dataset. However, this performance score relates to the assessment of RAPD and not necessarily the accuracy of pupil detection and tracking. In the results presented, we observe challenging conditions in the video data for which the pupil detection stage of the proposed pipeline fails to detect the pupil correctly. These challenging conditions include excessive blinking, motion blur from rapid eye movement, and incorrect pupil detection due to corneal reflection and its overlap onto the pupil.

A particular method that is able to detect the pupil accurately with a minimum number of feature points is the Starburst algorithm. This is useful for instances with excessive blinking, in which the pupil is regularly occluded by the subject's eyelids. Starburst algorithm also includes a significant pre-processing step for corneal reflection detection and removal [23], utilizing the brightness of the reflection for detection, and Gaussian smoothing to remove the reflection.

We also avoid using video data containing color, however, the inclusion of the color data would provide more sample points for which to assess RAPD. Therefore, color based image enhancement techniques such as color space manipulation [9] can be implemented to improve consistency in the color video data. The iris color has also been associated with pathological ocular complications [27], therefore, the addition of the color video data can improve the scope of future algorithmic development for ocular examination.

There is also a limitation seen in the pupil contour estimation using Circular Hough

Transform. In particular, some subjects present pupils that are non-circular. Therefore, a more flexible contour fitting algorithm can be implemented, such as ellipse fitting. This requires the inclusion of model-based selection of feature points, using models such as the random sample consensus (RANSAC) model [23], to minimize incorrect feature points.

In recent years, machine learning has been utilized in a variety of applications, especially in applications containing labelled data. The dataset introduced in this thesis contains pupillary light reflex video data, with corresponding medical ground truth data. Therefore, a possible research direction can include developing data-driven approaches for automated RAPD examination. We can utilize data-driven algorithms to learn inherent correlations in the video data, with respect to the corresponding ground truth labels available. This can improve the flexibility of the algorithm without requiring parameter tuning for new data.

The work presented in this thesis relates to the examination of RAPD. However, we can also apply the proposed algorithm for other applications that involve pupil information. For example, attention and interest can be assessed by the pupil size, in which the interest value of the visual activity is linked to the pupil size dilation during the activity [28]. The pupillary response can also reflect memory recollection and familiarity [29], which can aid in recuperation of patients recovering from memory loss. Therefore, the work presented in this research can be applied to other medical interests.

Appendices

APPENDIX A

POST-PROCESSED PUPIL SIZE MEASUREMENTS AND THEIR CORRESPONDING RAPD ASSESSMENT SCORES

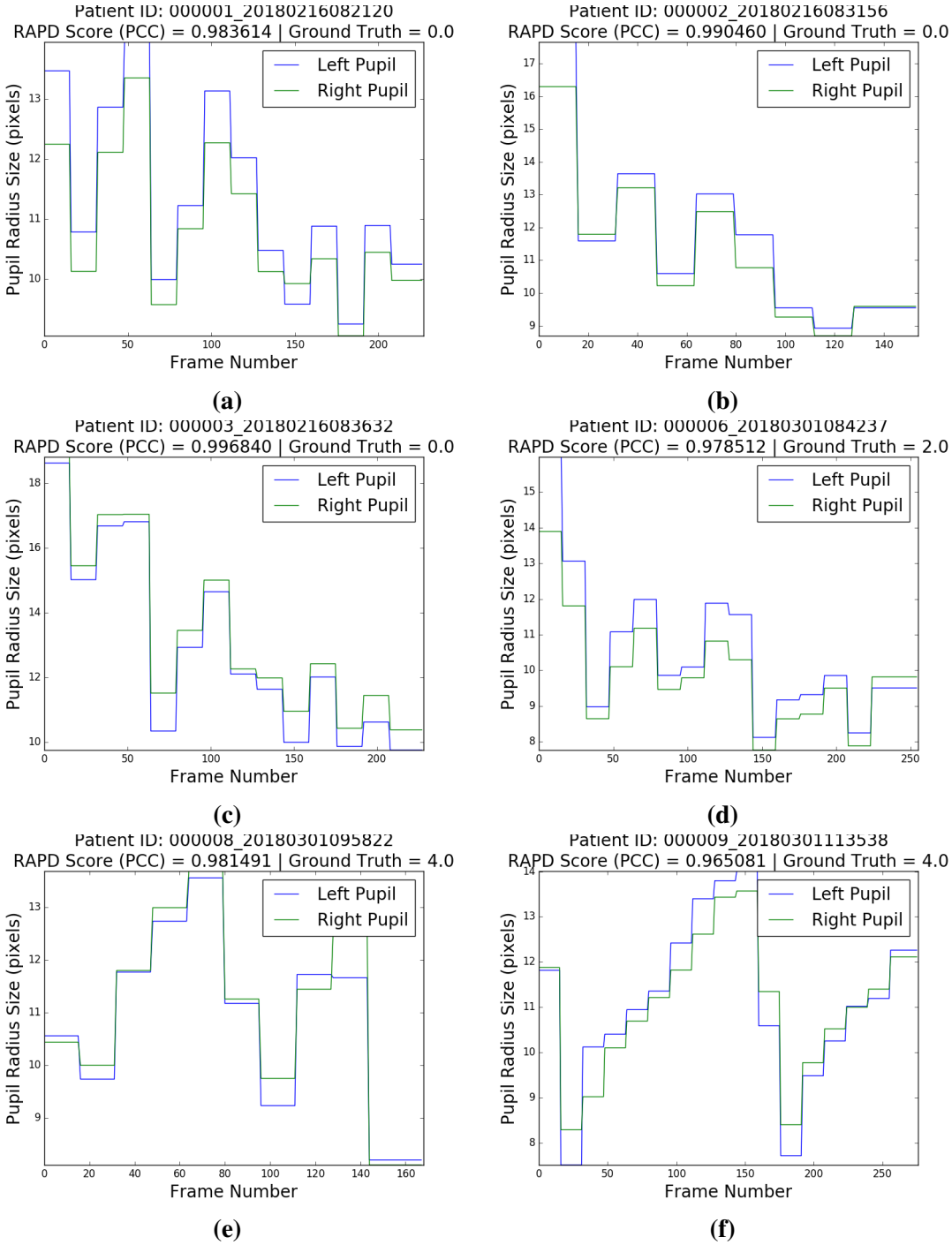


Figure 42: Pupil measurement data for subjects 1, 2, 3, 6, 8 and 9.

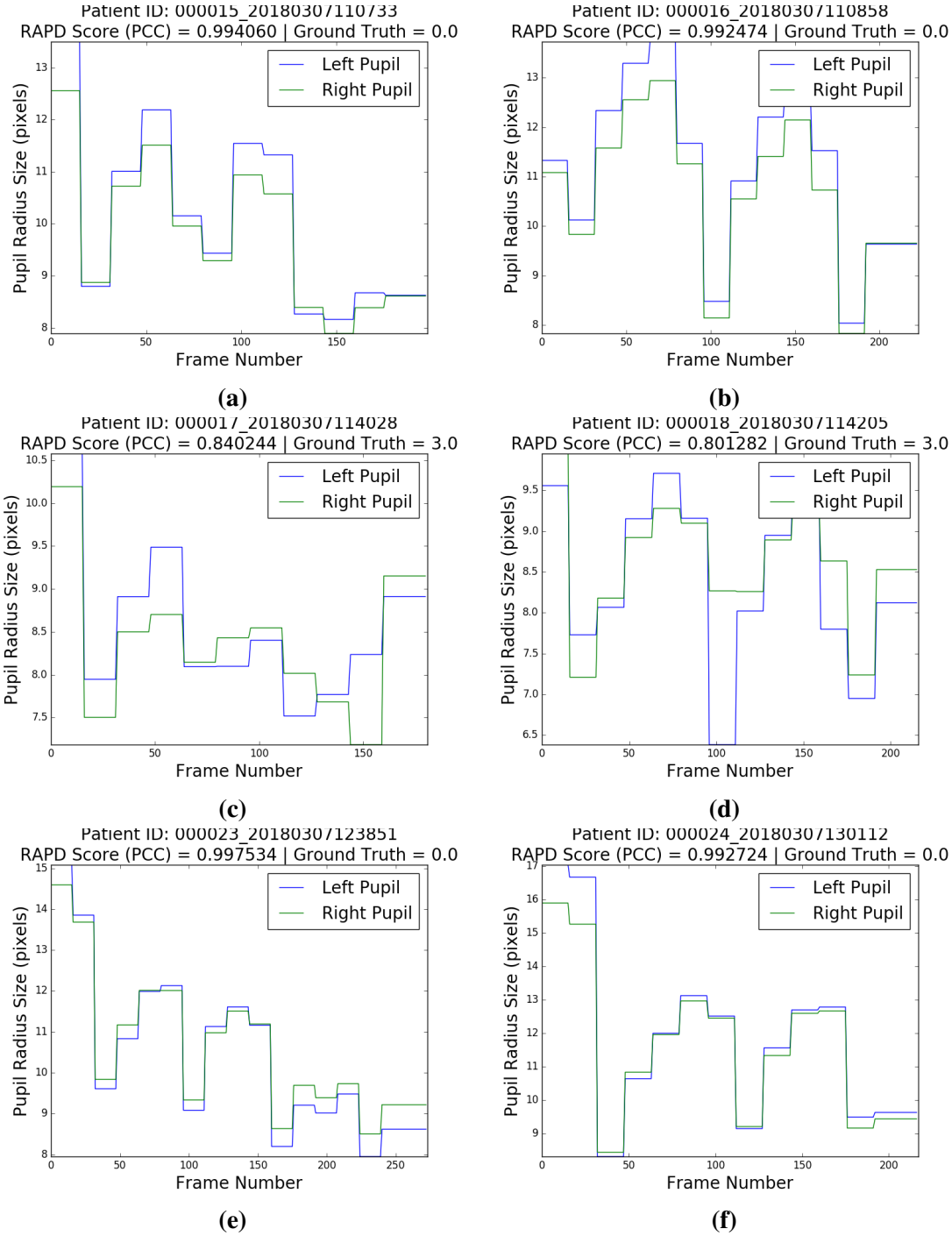


Figure 43: Pupil measurement data for subjects 15, 16, 17, 18, 23 and 24.

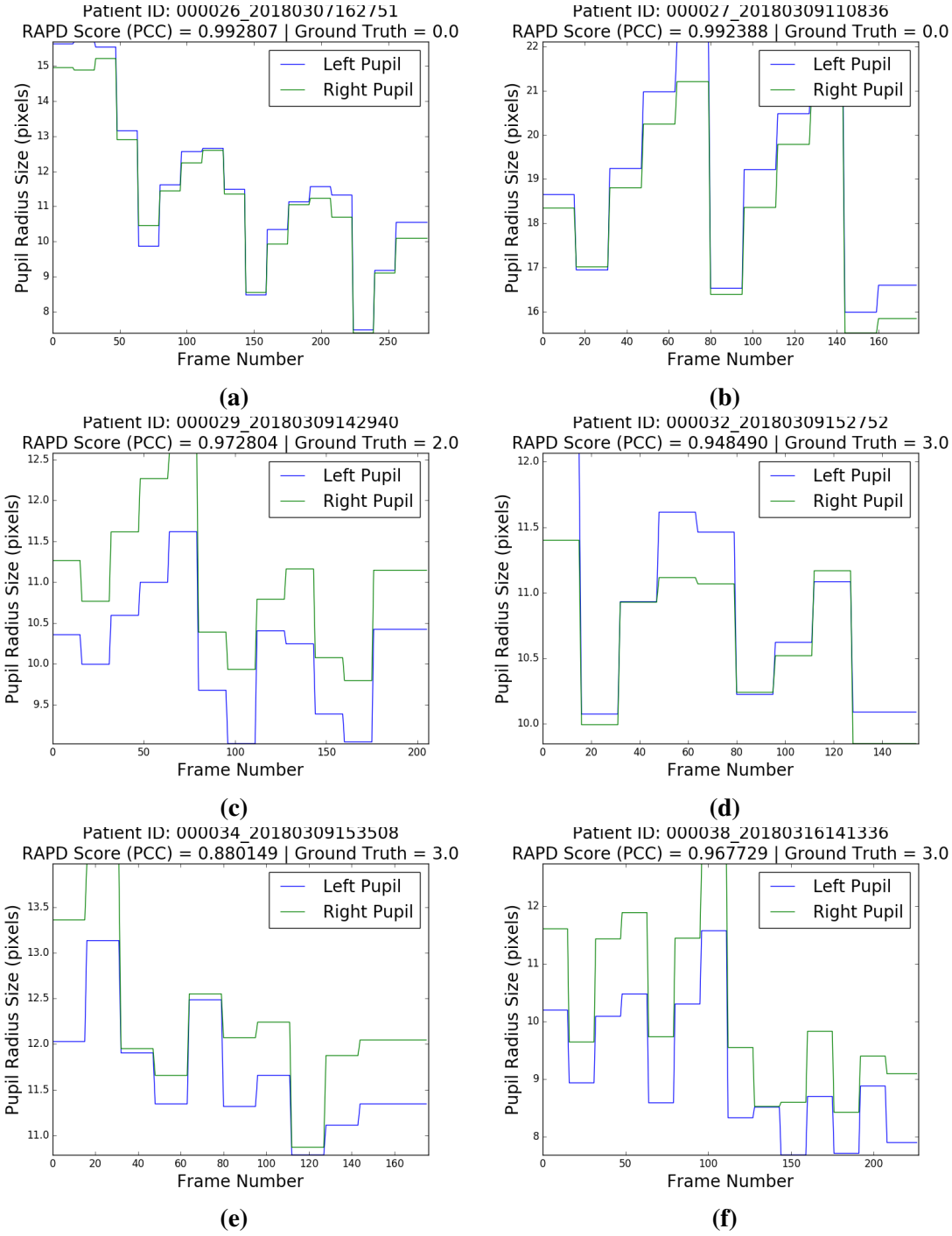


Figure 44: Pupil measurement data for subjects 26, 27, 29, 32, 34 and 38.

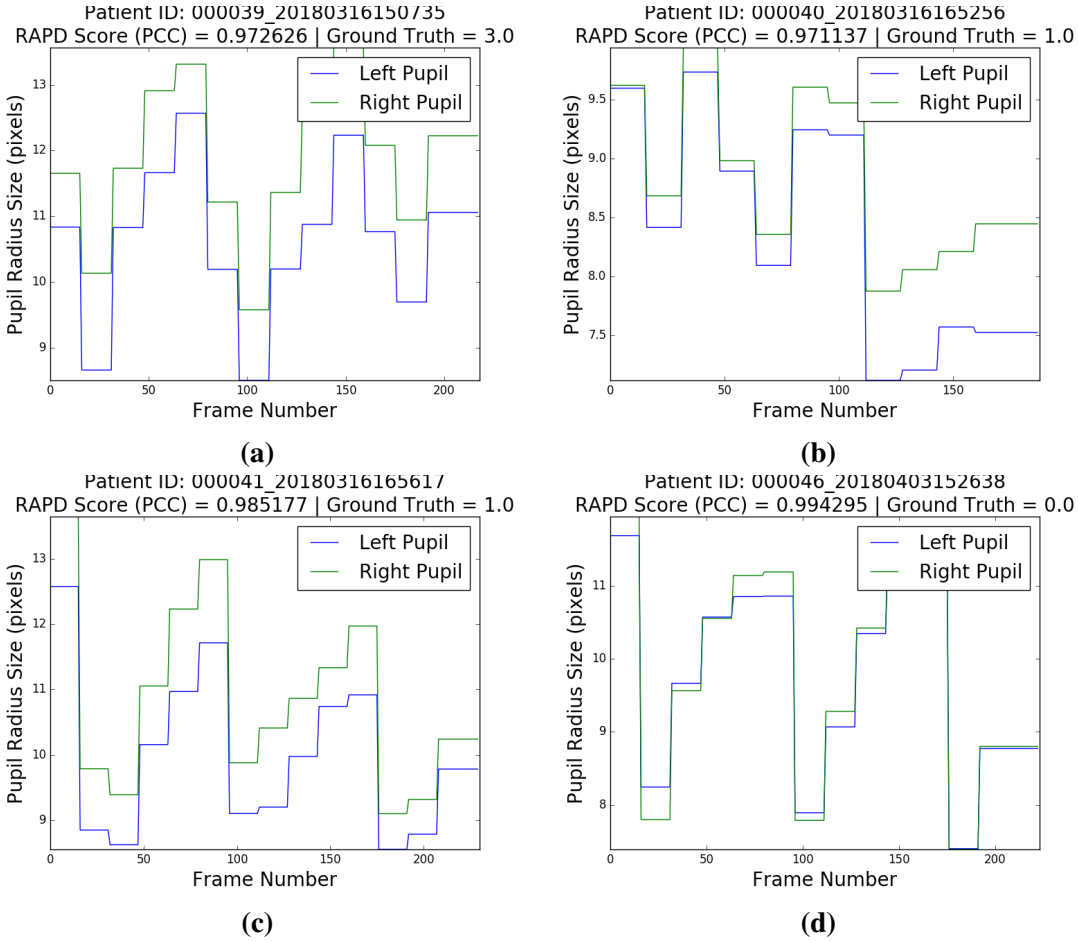


Figure 45: Pupil measurement data for subjects 39, 40, 41 and 46.

REFERENCES

[1] L. Omburo, S. Stutzman, C. Supnet, M. Choate, and D. M. Olson, "High variance in pupillary examination findings among postanesthesia care unit nurses," *Journal of PeriAnesthesia Nursing*, vol. 32, no. 3, pp. 219 –224, 2017.

[2] A. A. of Ophthalmology, "Eye Health Statistics," [Online]. Available: https://www.aao.org/newsroom/eye-health-statistics. [Accessed: 01-Apr-2018].

[3] Z. Tertel, "Analyzing ophthalmologys manpower issue," *Ophthalmology Management*, 2015. [Online]. Available: https://www.ophthalmologymanagement.com/issues/2015/january-2015 /analyzing-ophthalmology-8217;s-manpower-issue. [Accessed: 01-Apr-2018].

[4] R. N. I. of Blind People, "Workforce Census 2016," 2017. [Online]. Available: http://www.rnib.org.uk/nb-online/workforce-census-2016. [Accessed: 01-Apr-2018].

[5] StanfordMedicine, "Pupillary Responses," [Online]. Available: https://stanfordmedicine25.stanford.edu/the25/pupillary.html. [Accessed: 01-Jan-2018].

[6] C. B. Pate, "The proper procedure for testing pupils," [Online]. Available: http://optometrytimes.modernmedicine.com/optometrytimes/news/proper-procedure-testing-pupils. [Accessed: 01-Jan-2018].

[7] H. Gray, *Anatomy of the Human Body*. Lea and Febiger, 1918.

[8] M. G. Masi, L. Peretto, R. Tinarelli, and L. Rovati, "Measurement of the pupil diameter under different light stimula," *2009 IEEE Intrumentation Meas. Technol. Conf. I2MTC*, pp. 1652–1656, 2009.

[9] Y. Adachi, K. Konishi, M. Ozaki, and Y. Iwahori, "Development of an automatic measurement system of diameter of pupil - As an indicator of comprehension among web-based learners," *Procedia Comput. Sci.*, vol. 22, pp. 772–779, 2013.

[10] J. Chinni and H. V. Reddy, "Iris Recognition based on Pupil using Canny edge detection and K-Means Algorithm," vol. 2, no. 1, pp. 1–5, 2013.

[11] G. Xin, C. Ke, and H. Xiaoguang, "An improved Canny edge detection algorithm for color image," pp. 113–117, 2012.

[12] A. Jain, M. Gupta, S. N. Tazi, and Deepika, "Comparison of edge detectors," *2014 International Conference on Medical Imaging, m-Health and Emerging Communication Systems (MedCom), Greater Noida*, pp. 289–294, 2014.

[13] W. Yang, X. Wang, B. Moran, A. Wheaton, and N. Cooley, "Efficient registration of optical and infrared images via modified Sobel edging for plant canopy temperature estimation," *Comput. Electr. Eng.*, vol. 38, no. 5, pp. 1213–1221, 2012.

[14] M. B. Ahmad and T. Choi, "Local threshold and Boolean function based edge detection," *IEEE Trans. Consum. Electron.*, vol. 45, no. 1, pp. 674–679, 1999.

[15] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.

[16] L. Swirski, A. Bulling, and N. Dodgson, "Robust real-time pupil tracking in highly off-axis images," in *Proceedings of the Symposium on Eye Tracking Research and Applications*, ser. ETRA '12, Santa Barbara, California: ACM, 2012, pp. 173–176, ISBN: 978-1-4503-1221-9.

[17] M. Kassner, W. Patera, and A. Bulling, *Pupil: An open source platform for pervasive eye tracking and mobile gaze-based interaction*, 2014. eprint: `arXiv:1405.0006`.

[18] A. H. Javadi, Z. Hakimi, M. Barati, V. Walsh, and L. Tcheang, "SET: a pupil detection method using sinusoidal approximation," in *Frontiers in Neuroengineering*, 2015.

[19] W. Fuhl, T. Kubler, K. Sippel, W. Rosenstiel, and E. Kasneci, "Excuse: Robust pupil detection in real-world scenarios," in *Computer Analysis of Images and Patterns*, G. Azzopardi and N. Petkov, Eds., Cham: Springer International Publishing, 2015, pp. 39–51.

[20] W. Fuhl, T. C. Santini, T. Kubler, and E. Kasneci, "Else: Ellipse selection for robust pupil detection in real-world environments," in *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*, ser. ETRA '16, Charleston, South Carolina: ACM, 2016, pp. 123–130, ISBN: 978-1-4503-4125-7.

[21] N. Cherabit, F. Chelali, and A. Djeradi, "Circular Hough Transform for Iris localization," *Science and Technology*, vol. 2, no. 5, pp. 114–121, 2012.

[22] E. Rault, S. Aylward, and J. Jomier, "Automatic Quantification of Pupil Dilation Under Stress," *Computer-Aided Diagnosis and Display Lab , Department of Radiology The University of North Carolina*, pp. 249–252, 2004.

[23] D. Li, D. Winfield, and D. J. Parkhurst, "Starburst: A hybrid algorithm for video-based eye tracking combining feature-based and model-based approaches," *2005*

IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops, San Diego, CA, USA, vol. 2, no. 5, pp. 79–79, 2005.

[24] C. Kimme, D. Ballard, and J. Sklansky, "Find circles by an array of accumulators," *Comm. of the ACM*, vol. 18, no. 2, pp. 120–122, 1975.

[25] W. J. Ryan, D. L. Woodard, A. T. Duchowski, and S. T. Birchfield, "Adapting Starburst for Elliptical Iris Segmentation," *2008 IEEE Second International Conference on Biometrics: Theory, Applications and Systems, Arlington, VA*, pp. 1–7, 2008.

[26] F. Mokhayeri, S. Toosizadeh, and M. R. Akbarzadeh-T, "A Novel Approach for Pupil Diameter Measurement Based on Soft Computing Techniques," *2011 7th Iranian Conference on Machine Vision and Image Processing, Tehran*, pp. 1–5, 2011.

[27] H.-P. Sun, Y. Lin, and C.-W. Pan, "Iris color and associated pathological ocular complications: a review of epidemiologic studies," *International Journal of Ophthalmology*, vol. 7, no. 5, pp. 872–878, 2014.

[28] E. H. Hess and J. M. Polt, "Pupil Size as Related to Interest Value of Visual Stimuli," *Science*, vol. 132, no. 3423, pp. 349–350, 1960.

[29] A. Kafkas and D. Montaldi, "Familiarity and recollection produce distinct eye movement, pupil and medial temporal lobe responses when memory strength is matched," *Neuropsychologia*, vol. 50, no. 13, pp. 3080–3093, 2012.